3GPP2 C.S0018-D

Version 1.0

Date: January 25, 2010

Minimum Performance Specification for the Enhanced Variable Rate Codec, Speech Service Options 3, 68, 70, and 73 for Wideband Spread Spectrum Digital Systems

© 2010 3GPP2

3GPP2 and its Organizational Partners claim copyright in this document and individual Organizational Partners may copyright and issue documents or standards publications in individual Organizational Partner's name based on this document. Requests for reproduction of this document should be directed to the 3GPP2 Secretariat at secretariat@3gpp2.org. Requests to reproduce individual Organizational Partner's documents should be directed to that Organizational Partner. See www.3gpp2.org for more information.

REVISION HISTORY

Revision		Date
C.S0018-0 v1.0	Minimum Performance Specification for the Enhanced Variable Rate Codec, Speech Service Option 3 for Spread Spectrum Digital Systems	December 1999
C.S0018-B v1.0	Minimum Performance Specification for the Enhanced Variable Rate Codec, Speech Service Options 3 and 68 for Spread Spectrum Digital Systems	August 2007
C.S0018-C v1.0	Minimum Performance Specification for the Enhanced Variable Rate Codec, Speech Service Options 3, 68, and 70 for Spread Spectrum Digital Systems	December 2007
C.S0018-D v1.0	Minimum Performance Specification for the Enhanced Variable Rate Codec, Speech Service Options 3, 68, 70, and 73 for Spread Spectrum Digital Systems	January 25, 2010

FOREWORD

(This foreword is not part of the Standard.)

This document specifies the procedures to test implementations of EVRC-A, EVRC-B, EVRC-WB, or EVRC-NW compatible variable-rate speech codecs either by meeting the bit-exact implementation, or meeting recommended minimum performance requirements. The EVRC-A is the Service Option 3 (SO 3) speech codec, the EVRC-B is the Service Option 68 (SO 68) speech codec, the EVRC-WB is the Service Option 70 (SO 70) speech codec, and the EVRC-NW is the Service Option 73 (SO 73) speech codec.

REFERENCES

The following standards contain provisions which, through reference in this text, constitute provisions of this Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this Standard are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. ANSI, 3GPP2, TIA, and ITU-T maintain registers of currently valid national and international standards published by them.

NORMATIVE REFERENCES

1. 3GPP2 C.S0014-D v2.0, *Enhanced Variable Rate Codec, Speech Service Options 3, 68, 70, and 73 for Wideband Spread Spectrum Digital Systems*, January 2010.

1a. 3GPP2 C.R0014-C v1.0*, Software Distribution for Enhanced Variable Rate Codec, Speech Service Options 3, 68, and 70 for Wideband Spread Spectrum Digital Systems*. September 2007.

2. 3GPP2 C.S0009-0 v1.0, *Speech Service Option Standard for Wideband Spread Spectrum Systems*, December 1999.

3. 3GPP2 C.S0018-0 v1.0, *Minimum Performance Specification for the Enhanced Variable Rate Codec, Speech Service Option 3 for Spread Spectrum Digital Systems*, December 1999.

4. ANSI S1.4-1983 (R2006), *Sound Level Meters, Specification for*, March 2006.

5. ANSI S1.4A-1985 (R2006), *Sound Level Meters, Specifications for (Supplement to ANSI S1.4-1983)*, March 2006.

6. ITU-T Recommendation G.191, *Software Tools for Speech and Audio Coding Standardization*, September 2005.

6a. User's Group on Software Tools, ITU-T Software Tool Library 2005 User's Manual (Distributed with the software for STL2000), September 2005.

7. ITU-T Recommendation G.711, *Pulse code modulation (PCM) of voice frequencies*, November 1988.

8. ITU-T Recommendation G.728, *Coding of speech at 16 kbit/s using low-delay code excited linear prediction*, September 1992.

9. ITU-T Recommendation P.56, *Objective Measurement of Active Speech Level*, March 1993.

10. ITU-T Recommendation P.800, *Methods for Subjective Determination of Transmission Quality*. Annex B – Listening Tests – Absolute Category Rating (ACR), August 1996.

11. ITU-T Recommendation P.810, *Modulated Noise Reference Unit* (MNRU), February 1996.

12. ITU-T Recommendation P.830, *Methods for Objective and Subjective Assessment of Quality*. Annex – D Modified IRS Send and Receive Characteristics (MIRS), February 1996.

13. ITU-T Recommendation P.835, *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*, November 2003.

INFORMATIVE REFERENCES

14. 3GPP2 C.S0011-C, *Recommended Minimum Performance Standards for cdma2000 Spread Spectrum Mobile Stations*, March 2006.
15. 3GPP2 C.S0010-C, *Recommended Minimum Performance Standards for cdma2000 Spread Spectrum Base Stations*, March 2006.
16. TIA/EIA-95-B (R2004), *Mobile Station-Base Station Compatibility Standard for Wideband Spread Spectrum Cellular Systems,* October 2004.
17. IEEE 269-2002 (R2002), *Standard Method for Measuring Transmission Performance of Telephone Sets, Handsets, and Headsets,* April 2003.
18. IEEE STD 661-1979 (R2008), *IEEE Standard Method for Determining Objective Loudness Ratings of Telephone Connections,* September 2008.
19. ISO/IEC 9899-1999 (C2001/C2004), Programming Languages-C.
20. Dunnett, C.W., “A multiple comparison procedure for comparing several treatments with a control”, *Journal of the American Statistical Association*, vol.50, 1955, pp.1096-1121.
21. 3GPP TS 26.131, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Terminal acoustic characteristics for telephony; Requirements, V9.0.0, March 2009.

CONTENTS

1 Introduction 1-1
1.1 Scope 1-2
1.2 Definitions 1-3
1.3 Test Model for the Speech Codec 1-5
2 Codec Minimum Standards 2-1
2.1 Performance Testing for SO 3 2-1
2.1.1 Objective Performance Testing for SO 3 2-1
2.1.2 Subjective Performance Testing for SO 3 2-2
2.1.3 Source Speech Material for SO 3 Testing 2-7
2.1.4 Processing of Speech Material for SO 3 Testing 2-8
2.1.5 Randomization 2-10
2.1.6 Presentation 2-11
2.1.7 Listeners 2-11
2.1.8 Listening Test Procedures 2-11
2.1.9 Analysis of Results 2-12
2.1.10 Minimum Subjective Requirement 2-13
2.1.11 Expected Results for Reference Conditions 2-15
2.2 Performance Testing for SO 68 2-17
2.2.1 Objective Performance Testing for SO 68 2-17
2.2.2 Subjective Performance Testing for SO 68 2-19
2.2.3 Speech Material for SO 68 Testing 2-23
2.2.4 Processing of Speech Material for SO 68 Testing 2-24
2.2.5 Randomization 2-27
2.2.6 Presentation 2-29
2.2.7 Listeners 2-29
2.2.8 Listening Test Procedures 2-29
2.2.9 Analysis of Results 2-33
2.2.10 Expected Results for Reference Conditions 2-34
2.3 Performance Testing for SO 70 2-35
2.3.1 Objective Performance Testing for SO 70 2-35
2.3.2 Subjective Performance Testing for SO 70 2-37

2.3.3 Speech Material for SO 70 Testing 2-52
2.3.4 Processing of Speech Material for SO 70 Testing 2-52
2.3.5 Randomization 2-56
2.3.6 Presentation 2-58
2.3.7 Listeners 2-58
2.3.8 Listening Test Procedures 2-58
2.3.9 Analysis of Results 2-62
2.3.10 Expected Results for Reference Conditions 2-63
2.4 Performance Testing for SO 73 2-64
2.4.1 Objective Performance Testing for SO 73 2-64
2.4.2 Subjective Performance Testing for SO 73 2-66
2.4.3 Speech Material for SO 73 Testing 2-78
2.4.4 Processing of Speech Material for SO 73 Testing 2-78
2.4.5 Randomization 2-81
2.4.6 Presentation 2-83
2.4.7 Listeners 2-83
2.4.8 Listening Test Procedures 2-83
2.4.9 Analysis of Results 2-87
2.4.10 Expected Results for Reference Conditions 2-88
3 Codec Standard Test Conditions 3-1
3.1 Specific Standard Test Conditions for SO 3 3-2
3.1.1 Audio Path and Calibration for SO 3 3-2
3.1.2 Standard Software Test Tools for SO 3 3-3
3.1.3 Master Codec for SO 3 3-5
3.1.4 Fixed-Point Bit-Exact Codec for SO 3 3-7
3.2 Specific Standard Test Conditions for SO 68 3-13
3.2.1 Audio Path and Calibration for SO 68 3-13
3.2.2 Standard Software Test Tools for SO 68 3-14
3.2.3 Master Codec for SO 68 3-15
3.2.4 Fixed-Point Bit-Exact Codec for SO 68 3-16
3.3 Specific Standard Test Conditions for SO 70 3-20
3.3.1 Audio Path and Calibration for SO 70 3-20
3.3.2 Software Test Tools for SO 70 3-21

3.3.3 Master Codec for SO 70 3-22
3.3.4 Fixed-Point Bit-Exact Codec for SO 70 3-24
3.4 Specific Standard Test Conditions for SO 73 3-34
3.4.1 Audio Path and Calibration for SO 73 3-34
3.4.2 Software Test Tools for SO 73 3-35
3.4.3 Master Codec for SO 73 3-36
3.4.4 Fixed-Point Bit-Exact Codec for SO 73 3-38
4 CONTENTS OF SOFTWARE DISTRIBUTION 4-1
5 DUNNETT'S TEST 5-1
5.1 Stage 1 – Analysis of Variance 5-1
5.2 Stage 2 – Dunnett's Multiple Means Test — Test CC's vs. the Reference CC 5-2
6 Processing BLOCKs FOR So 68, SO 70, and SO 73 6-1
6.1 Nominal Level, and Noise Processing 6-1
6.2 FER Processing 6-1
6.3 Low-level, and Signaling Processing 6-1
6.4 High level Processing 6-1

LIST OF FIGURES

Figure 1.3-1 Test Model 1-6
Figure 2.1.8-1 Instructions for Listeners 2-12
Figure 2.1.11-1 MOS versus MNRU 2-16
Figure 2.2.8.1-1 Instructions for Listeners 2-30
Figure 2.2.10.1-1 MOS versus MNRU 2-34
Figure 2.2.10.2-1 P.835 Score Profiles for Reference Conditions 2-35
Figure 2.3.2-1 SO 70 Subjective test suite decision flowchart 2-38
Figure 2.3.8.1-1 Instructions for Listeners 2-59
Figure 2.3.10.1-1 Typical Plot of MOS versus MNRU 2-63
Figure 2.3.10.2-1 Typical P.835 Score Profiles for Reference Conditions 2-64
Figure 2.4.2-1 SO 73 Subjective test suite decision flowchart 2-67
Figure 2.4.8.1-1 Instructions for Listeners 2-84
Figure 2.4.10.1-1 Typical Plot of MOS versus MNRU 2-88
Figure 2.4.10.2-1 Typical P.835 Score Profiles for Reference Conditions 2-89
Figure 3-1 Basic Test Equipment 3-1
Figure 3-2 Subjective Testing Equipment Configuration 3-2
Figure 3.3.2.4-1 SO 70 ITU-T P.311/P.341 Transmit Mask and Filter responses 3-22
Figure 3.3.4.5-1 SO 70 Fixed-point bit-exact test suite decision flowchart 3-26
Figure 3.4.2.4-1 SO 73 ITU-T P.311/P.341 Transmit Mask and Filter responses 3-36
Figure 3.4.4.5-1 SO 73 Fixed-point bit-exact test suite decision flowchart 3-40

LIST OF TABLES

Table 2.1.2.3.1-1 SO 3 Listening Experiment I Conditions 2-3
Table 2.1.2.3.1-2 SO 3 Listening Experiment I Design 2-4
Table 2.1.2.3.2-1 SO 3 Listening Experiment II Conditions 2-5
Table 2.1.2.3.2-2 SO 3 Listening Experiment II Design 2-5
Table 2.1.2.3.3-1 Numerical Parameters for SO 3 Listening Experiments 2-7
Table 2.1.10-1 Multipliers for Equations 2.1.10-4 and 2.1.10-5 2-15
Table 2.2.1.1.1-1 Target ADR vs Capacity Operating Point 2-17
Table 2.2.2.3.1-1 SO 68 Listening Experiment I Test Parameters 2-19
Table 2.2.2.3.1-2 SO 68 Listening Experiment I Test Conditions 2-20
Table 2.2.2.3.2-1 SO 68 Listening Experiment II Test Parameters 2-21
Table 2.2.2.3.2-2 SO 68 Listening Experiment II Test Conditions 2-22
Table 2.2.2.3.3-1 Numerical Parameters for the SO 68 Listening Experiments 2-23
Table 2.2.4.5-1 Cutting Points for the astrip Software Tool for the Experiment I ACR Test 2-26
Table 2.2.4.5-2 Cutting Points for the *astrip* Software Tool for the Experiment II P.835 Test 2-27
Table 2.2.4.5-3 Composition of the Sentence-Triad Samples for the Experiment II P.835 Test 2-27
Table 2.2.5-1 Example Randomization for the Experiment I ACR Test 2-28
Table 2.3.1.1.1-1 Target ADR vs Capacity Operating Point 2-36
Table 2.3.2-1 Test Suites for SO 70 compliance 2-39
Table 2.3.2-2 Experiments for SO 70 compliance 2-39
Table 2.3.2.3.1-1 SO 70 Listening Experiment 1 Test Parameters 2-40
Table 2.3.2.3.1-2 SO 70 Listening Experiment 1 Test Conditions 2-41
Table 2.3.2.3.2-1 SO 70 Listening Experiment 2 Test Parameters 2-42
Table 2.3.2.3.2-2 SO 70 Listening Experiment 2 Test Conditions. 2-43
Table 2.3.2.3.3-1 SO 70 Listening Experiment 3 Test Parameters 2-44
Table 2.3.2.3.3-2 SO 70 Listening Experiment 3 Test Conditions. 2-44
Table 2.3.2.3.4-1 SO 70 Listening Experiment 4 Test Parameters 2-46
Table 2.3.2.3.4-2 SO 70 Listening Experiment 4 Test Conditions. 2-46
Table 2.3.2.3.5-1 SO 70 Listening Experiment 5 Test Parameters 2-47
Table 2.3.2.3.5-2 SO 70 Listening Experiment 5 Test Conditions. 2-48
Table 2.3.2.3.6-1 SO 70 Listening Experiment 6 Test Parameters 2-49
Table 2.3.2.3.6-2 SO 70 Listening Experiment 6 Test Conditions. 2-49

Table 2.3.2.3.7-1 SO 70 Listening Experiment 7 Test Parameters 2-50
Table 2.3.2.3.7-2 SO 70 Listening Experiment 7 Test Conditions. 2-50
Table 2.3.2.3.8-1 SO 70 Listening Experiment 8 Test Parameters 2-51
Table 2.3.2.3.8-2 SO 70 Listening Experiment 8 Test Conditions. 2-51
Table 2.3.2.3.9-1 Numerical Parameters for the SO 70 Listening Experiments 2-52
Table 2.3.4.5-1 Cutting Points for the astrip Software Tool for the SO 70 Experiments 1, 3, and 5 ACR Test 2-55
Table 2.3.4.5-2 Cutting Points for the *astrip* Software Tool for the SO 70 Experiments 2, 4, and 6 P.835 Test 2-56
Table 2.3.4.5-3 Composition of the Sentence-Triad Samples for the Experiments 2, 4, and 6 P.835 Test 2-56
Table 2.3.5-1 Example Randomization for the Experiments 1, 3, and 5 ACR Test 2-57
Table 2.4.1.1.1-1 Target ADR vs Capacity Operating Point 2-65
Table 2.4.2-1 Test Suites for SO 73 compliance 2-67
Table 2.4.2-2 Experiments for SO 73 compliance 2-67
Table 2.4.2.3.1-1 SO 73 Listening Experiment 1 Test Parameters 2-68
Table 2.4.2.3.1-2 SO 73 Listening Experiment 1 Test Conditions 2-69
Table 2.4.2.3.2-1 SO 74 Listening Experiment 2 Test Parameters 2-70
Table 2.4.2.3.2-2 SO 73 Listening Experiment 2 Test Conditions. 2-70
Table 2.4.2.3.3-1 SO 73 Listening Experiment 3 Test Parameters 2-71
Table 2.4.2.3.3-2 SO 73 Listening Experiment 3 Test Conditions. 2-72
Table 2.4.2.3.4-1 SO 74 Listening Experiment 4 Test Parameters 2-73
Table 2.4.2.3.4-2 SO 73 Listening Experiment 4 Test Conditions. 2-74
Table 2.4.2.3.5-1 SO 73 Listening Experiment 5 Test Parameters 2-75
Table 2.4.2.3.5-2 SO 73 Listening Experiment 5 Test Conditions. 2-76
Table 2.4.2.3.6-1 SO 73 Listening Experiment 6 Test Parameters 2-76
Table 2.4.2.3.6-2 SO 73 Listening Experiment 6 Test Conditions. 2-77
Table 2.4.2.3.7-1 Numerical Parameters for the SO 73 Listening Experiments 2-77
Table 2.4.4.5-1 Cutting Points for the astrip Software Tool for the SO 73 Experiments 1 and 3 ACR Test 2-80
Table 2.4.4.5-2 Cutting Points for the *astrip* Software Tool for the SO 73 Experiments 2 and 4 P.835 Test 2-81
Table 2.4.4.5-3 Composition of the Sentence-Triad Samples for the Experiments 2 and 4 P.835 Test 2-81
Table 2.4.5-1 Example Randomization for the Experiments 1 and 3 ACR Test 2-82

Table 3.1.3.3-1 Packet File Structure From Master Codec/Channel Error Model 3-6
Table 3.1.4.6.2-1 Source and Bit-exact Default Mode Test Vector Files 3-11
Table 3.1.4.6.2-2 Source and Bit-exact Rate-1/2 Max Test Vector Files 3-12
Table 3.1.4.6.2-3 Source and Bit-exact Full Rate Only Test Vector Files 3-12
Table 3.1.4.6.2-4 Decoder Output Test Vector Files 3-13
Table 3.2.3.3-1 Packet File Structure From Master Codec/Channel Error Model 3-16
Table 3.2.4.5-1 SO 68 Encoder Bit-exact Test Conditions 3-18
Table 3.2.4.5-2 SO 68 Decoder Bit-exact Test Conditions 3-19
Table 3.3.3.3-1 Packet File Structure From Master Codec/Channel Error Model 3-23
Table 3.3.4.5-1 Test Suites of input test vectors for SO 70 compliance 3-27
Table 3.3.4.5-2 SO 70 Encoder Suite A Bit-exact Test Conditions 3-27
Table 3.3.4.5-3 SO 70 Suite A Decoder Bit-exact Test Conditions 3-28
Table 3.3.4.5-4 SO 70 Encoder Suite B Bit-exact Test Conditions 3-28
Table 3.3.4.5-5 SO 70 Suite B Decoder Bit-exact Test Conditions 3-29
Table 3.3.4.5-6 SO 70 Encoder Suite C Bit-exact Test Conditions 3-31
Table 3.3.4.5-7 SO 70 Suite C Decoder Bit-exact Test Conditions 3-31
Table 3.3.4.5-8 SO 70 Encoder Suite D Bit-exact Test Conditions 3-32
Table 3.3.4.5-9 SO 70 Suite D Decoder Bit-exact Test Conditions 3-32
Table 3.4.3.3-1 Packet File Structure from Master Codec/Channel Error Model 3-37
Table 3.4.4.5-1 Test Suites of input test vectors for SO 73 compliance 3-40
Table 3.4.4.5-2 SO 73 Encoder Suite A Bit-exact Test Conditions 3-41
Table 3.4.4.5-3 SO 73 Suite A Decoder Bit-exact Test Conditions 3-42
Table 3.4.4.5-4 SO 73 Encoder Suite B Bit-exact Test Conditions 3-43
Table 3.4.4.5-5 SO 73 Suite B Decoder Bit-exact Test Conditions 3-44
Table 4-1 Description of EVRC-A Software Distribution Contents 4-1
Table 4-2 Description of EVRC-B Software Distribution Contents 4-2
Table 4-3 Description of EVRC-WB Software Distribution Contents 4-2
Table 4-4 Description of EVRC-NW Software Distribution Contents 4-3
Table 5.1-1 Variance Source Table for the ANOVA 5-2

1 INTRODUCTION

This standard details definitions, methods of measurement, verification of bit-exactness, and minimum performance characteristics of the EVRC-A, EVRC-B, EVRC-WB, and EVRC-NW enhanced variable-rate speech codecs for digital cellular spread spectrum mobile stations and base stations, specified in [1][1]. This standard shares the purpose of [14] and [15]. This is to ensure that a mobile station can obtain service in any cellular system that meets the compatibility requirements of [16].

This standard consists of this document and an associated software distribution. The Software Distribution contains:

- Audio source material
- Clear channel packets produced from the master codec
- Impaired channel packets produced from the master codec and degraded by a channel model simulation
- Output audio files produced from the master encoded packets decoded by the master decoder
- Calibration source material
- C/C++ language source files for the compilation of bit-exact fixed-point codec
- C/C++ language source files for a number of software data analysis tools
- Modulated Noise Reference Unit (MNRU) reference files
- Input and output vectors for bit-exact testing

An overview of the contents and formats of the software distribution is given in Section 4 of this document.

The EVRC-A, EVRC-B, EVRC-WB, and EVRC-NW enhanced variable-rate speech codecs (collectively referred to as EVRC) are intended to be used at mobile stations at compatible base stations in the cellular service. This statement is not intended to preclude implementations in which codecs are placed at a Mobile Switching Center or elsewhere within the cellular system. Indeed, some mobile-to-mobile calls, however routed, may not require the implementation of a codec on the fixed side of the cellular system at all. This standard is meant to define both verifications of bit-exact implementations and the recommended minimum performance requirements of EVRC-compatible variable-rate codecs, no matter where or how they are implemented in the cellular service.

Although the basic purpose of cellular telecommunications has been voice communication, evolving usages (for example, data) may allow the omission of some of the features specified herein provided that system compatibility is not compromised.

[1]Numbers in brackets, [N], refer to the reference document numbers. For example, [1] refers reference 1 in the reference list.

This standard concentrates specifically on the EVRC, whether implemented at the mobile station or the base station or elsewhere in the cellular system. This standard covers the operation of this component only to the extent that compatibility with the specific EVRC-compatible variable-rate codec is ensured.

1.1 Scope

This document specifies the procedures to test implementations of EVRC-A, EVRC-B, EVRC-WB, or EVRC-NW compatible variable-rate speech codecs either by meeting the bit-exact implementation, or meeting recommended minimum performance requirements. The EVRC-A is the Service Option 3 (SO 3) speech codec, the EVRC-B is the Service Option 68 (SO 68) speech codec, the EVRC-WB is the Service Option 70 (SO 70) speech codec, and the EVRC-NW is the Service Option 73 (SO 73) speech codec, all described in [1]. The procedures specified in this document for the SO 3 speech codec are fully consistent with those contained in [3]. The SO 3 speech codec is used to digitally encode the speech signal for transmission at a variable data rate of 8550 bps, 4000 bps, or 800 bps. The SO 68 speech codec is used to digitally encode the speech signal for transmission at a variable data rate of 8550 bps, 4000 bps, 2000 bps, or 800 bps. The SO 70 speech codec is used to digitally encode the speech signal for transmission at a variable data rate of 8550 bps, 4000 bps, or 800 bps. The SO 73 speech codec is used to digitally encode the speech signal for transmission at a variable data rate of 8550 bps, 4000 bps, 2000 bps, or 800 bps.

Like some other speech coding standards, this standard provides a bit-exact method of verifying the test codec for minimum performance. In this optional procedure, a given set of test vectors are input to the test codec and the output vectors from the test codec must be bit-exact with the output vectors given in the software distribution which is associated with this standard. If they are bit-exact, the test codec passes the minimum performance requirement and no further testing is required. The bit-exact mode of testing, however, is only applicable to codecs whose design conforms in all respects to the algorithmic description of the specific EVRC service option, including the noise suppression, rate determination and post-filter components.

Should the candidate EVRC differ in any of these components, the test codec shall be tested using the objective and subjective tests prescribed by this standard. That is, EVRC-compliance of a "test codec" can be achieved by either:

- Complying with Sections 2.1.1 and 2.1.2 (SO 3), or Sections 2.2.1 and 2.2.2 (SO 68), or Sections 2.3.1 and 2.3.2 (SO 70), or Section 2.4.1 and 2.4.2 (SO 73), and demonstrating bit-exactness according to the procedure described in Section 3.1.4 (SO 3), or Section 3.2.4 (SO 68), or Section 3.3.4 (SO 70), or Section 3.4.4 (SO 73), respectively.
- Following the objective and subjective testing procedures set forth in Sections 2.1.1 and 2.1.2, or Sections 2.2.1 and 2.2.2, or Sections 2.3.1 and 2.3.2, or Section 2.4.1 and 2.4.2 (SO 73) of this standard.

With the exception of Sections 3.1.4, 3.2.4, 3.3.4, and 3.4.4, the remaining text applies only to implementations that do not satisfy the requirement for bit-exactness.

Testing the codec is based on two classes of procedures: objective tests and subjective tests. In the event that the test codec fails any of the objective or subjective tests, the test codec fails the compliance test. Objective tests are based upon actual measurements from the speech codec function. Subjective tests are based on listening tests to judge overall speech quality. The minimum

subjective requirement for the test codec is based upon the ability of the test codec to demonstrate performance equivalent to or better than that of the specific EVRC floating-point bit-exact codec within a fixed allowable statistical error.

The purpose of the testing is not only to ensure adequate performance between one manufacturer's encoder and decoder but also that this level of performance is maintained with operation between any pairing of manufacturers' encoders and decoders. This interoperability issue is a serious one. Any variation in implementing the exact standard must be avoided if it cannot be ensured that minimum performance levels are met when interoperating with all other manufacturers' equipment meeting the standard. This standard provides a means for measuring performance levels while trying to ensure proper interoperation with other manufacturers' equipment.

The issue of interoperation can only be definitively answered by testing all combinations of encoder/decoder pairings. With the number of equipment manufacturers expected to supply equipment, this becomes a prohibitive task; therefore, the objective and subjective tests rely upon the use of a "master codec". The master codec is defined as the floating-point implementation of specific EVRC written in the C programming language. The master codec software which is described in Section 3.1.3 (SO 3), Section 3.2.3 (SO 68), Section 3.3.3 (SO 70), or Section 3.4.3 (SO 73) is used as part of the interoperability testing.

1.2 Definitions

Base Station - A station in the Domestic Public Cellular Radio Telecommunications Service, other than a mobile station, used for radio communications with mobile stations.

Bit-Exact - A test procedure for codecs by which a set of prescribed vectors are input to the test codecs, and output vectors from the codecs correspond exactly bit-for-bit with output vectors prescribed by this standard.

CELP - Code Excited Linear Predictive Coding. This technique uses codebooks to vector quantize the excitation (residual) signal of a Linear Predictive Codec (LPC).

Circum-aural Headphones - Headphones that surround and cover the entire ear.

Codec - The combination of an encoder and decoder in series (encoder/decoder).

Compand - The process of compressing and expanding a signal. In this text, the process is described in terms of μ-Law PCM [7].

dB - Normally taken to be defined as: X dB = $20 \cdot \log_{10}(x)$. In the context of digitized speech, the unit dB is used to represent the average power level of a speech signal with respect to full scale. For the purposes of this document, "full scale" is defined as the maximum sinusoidal input level which does not result in clipping, where 0 dB corresponds to the output level, measured according to [9], for a full scale 1-kHz sinusoidal input. This corresponds to a digitally referenced input level of –3 dBov, and [7] defined tone level of +3.17 dBm0. Nominal input speech level is defined to be approximately 22 dB below this reference tone level, and is equivalent to –25 dBov, or –19 dBm0. For 16 bit signed integers, a sine wave with a peak amplitude of 32768 corresponds to 0 dB, according to this definition. Because a sine wave with amplitude A has a RMS value of $A/\sqrt{2}$, the level in dB of a voice active segment of speech $\{x(n),\ldots,x(n+N-1)\}$ quantized with 16-bit two's complement linear data spanning [-32768, 32767], is given by:

$$XdB = 10\log_{10}\left(\frac{2}{32768^2 N}\sum_{i=n}^{n+N-1} x^2(i)\right).$$

dBA - A-weighted sound pressure level expressed in decibels obtained by the use of a metering characteristic and the weighting A, specified in [4] and [5].

dBm0 - Power relative to 0 transmission level point (TLP). [7] specifies a theoretical load capacity with a full scale sine wave to be +3.17 dBm0 for μ-law PCM coding and +3.14 dBm0 for A-Law PCM coding.

dBPa - Sound level with respect to one Pascal, 20 $\log_{10}$ (Pressure/1 Pa).

dB SPL - Sound Pressure Level in decibels with respect to 0.002 dynes/cm^2, 20 $\log_{10}$ (Pressure/0.002 dynes/cm^2). dBPa is preferred.

Decoder - A device for the translation of a signal from a digital representation into an analog format. For the purposes of this standard, a device compatible with a specific EVRC implementation.

Encoder - A device for the coding of a signal into a digital representation. For the purpose of this standard, a device compatible with a specific EVRC implementation.

FER - Frame Error Rate equals the number of full rate frames received in error divided by the total number of transmitted frames.

IRS - Intermediate Reference System [12].

MGW – Media Gateway

MIRS – Modified Intermediate Reference System [12].

MNRU - Modulated Noise Reference Unit. A procedure to add speech correlated noise to a speech signal in order to produce distortions that are subjectively similar to that produced by logarithmically companded PCM systems. The amount of noise is expressed as a signal-to-noise ratio value in dB, and is usually referred to as dBQ [11].

Mobile Station - A station in the Domestic Public Cellular Radio Telecommunications Service. It is assumed that mobile stations include portable tranceivers (for example, hand-held personal tranceivers) and tranceivers installed in vehicles.

MOS - Mean Opinion Score. The result of a subjective test based on an absolute category rating (ACR), where listeners associate a quality adjective with the speech samples to which they are listening. These subjective ratings are transferred to a numerical scale, and the arithmetic mean is the resulting MOS number [10].

Rates for SO 3 - The allowable traffic frame rates for SO 3: Rate 1 frames use the 9600 bps rate, Rate ½ frames use the 4800 bps rate, Rate ¼ frames use the 2400 bps rate, and Rate 1/8 frames use the 1200 bps rate. The allowable speech encoding frame rates for SO 3: Rate 1 frames use the 8550 bps rate, Rate ½ frames use the 4000 bps rate, Rate ¼ frames are not used in Service Option 3, and Rate 1/8 frames use the 800 bps rate.

Rates for SO 68 - The allowable traffic frame rates for SO 68: Rate 1 frames use the 9600 bps rate, Rate ½ frames use the 4800 bps rate, Rate ¼ frames use the 2400 bps rate, and Rate 1/8 frames use the 1200 bps rate. The allowable speech encoding frame rates for SO 68: Rate 1 frames use the

8550 bps rate, Rate ½ frames use the 4000 bps rate, Rate ¼ frames use the 2000 bps rate, and Rate 1/8 frames use the 800 bps rate.

Rates for SO 70 - The allowable traffic frame rates for SO 70: Rate 1 frames use the 9600 bps rate, Rate ½ frames use the 4800 bps rate, and Rate 1/8 frames use the 1200 bps rate. The allowable speech encoding frame rates for SO 70: Rate 1 frames use the 8550 bps rate, Rate ½ frames use the 4000 bps rate, and Rate 1/8 frames use the 800 bps rate.

Rates for SO 73 - The allowable traffic frame rates for SO 73: Rate 1 frames use the 9600 bps rate, Rate ½ frames use the 4800 bps rate, Rate ¼ frames use the 2400 bps rate, and Rate 1/8 frames use the 1200 bps rate. The allowable speech encoding frame rates for SO 73: Rate 1 frames use the 8550 bps rate, Rate ½ frames use the 4000 bps rate, Rate ¼ frames use the 2000 bps rate, and Rate 1/8 frames use the 800 bps rate.

ROLR - Receive Objective Loudness Rating: a measure of receive audio sensitivity. ROLR is a frequency-weighted ratio of the line voltage input signal to a reference encoder to the acoustic output of the receiver. [17] defines the measurement of sensitivity, and [18] defines the calculation of objective loudness rating.

Supra-aural Headphones - Headphones that cover but do not surround the entire ear.

$\mathbf{T_{max}}$ - The maximum undistorted sinusoidal level that can be transmitted through the interfaces between the EVRC and the PCM-based network. This is taken to be a reference level of +3.17 dBm0.

1.3 Test Model for the Speech Codec

For the purposes of this standard, a speech encoder is a process that transforms a stream of binary data samples of speech into an intermediate low bit-rate parameterized representation. As mentioned elsewhere in this document, the reference method for the performance of this process is given in [1]. This process may be implemented in real-time as a software program or otherwise at the discretion of the manufacturer.

Likewise, a speech decoder is a process that transforms the intermediate low bit-rate parameterized representation of speech (given [1]) back into a stream of binary data samples suitable for input to a digital-to-analog converter followed by an electro-acoustic transducer.

The test model compares the output streams of the test encoder and/or decoder to those of a master encoder or decoder when driven by the same input stream. Figure 1.3-1 shows how the various combinations of outputs are generated. Various test conditions will dictate the specific source material and the functions of the gain blocks, the frame error model block, and the external rate control.

The input stream for an encoder is a sequence of 16-bit linear binary 2's complement samples of speech source material. The speech can be clean (no background noise) or can have background noise added, depending on the condition being tested. The source is passed through the gain block, which can amplify or attenuate the signal depending on the condition being tested. This signal is then processed by both the master and test encoders, with the ability to control the maximum packet rate externally. The output of the test encoder for a given rate must conform to the packet files formats specified in [1]. The master encoded speech packets can be presented to a frame error model which

simulates packet loss over a CDMA air interface. The (potentially corrupted) encoded speech packets from the master and test encoders are then used as inputs to each of the master and test decoders, forming four combinations of decoded outputs. The four output combinations are master encode/master decode, test encode/master decode, master encode/test decode, and test encode/test decode, or more simply: M/M, T/M, M/T, and T/T respectively. The decoded speech material is then appropriately gain adjusted (inversely to input gain) and formatted (μ-Law PCM for SO 3 and 16-bit linear PCM for SO 68, SO 70, and SO 73) to form the final outputs. The representation of output speech is the same as that for input speech material.

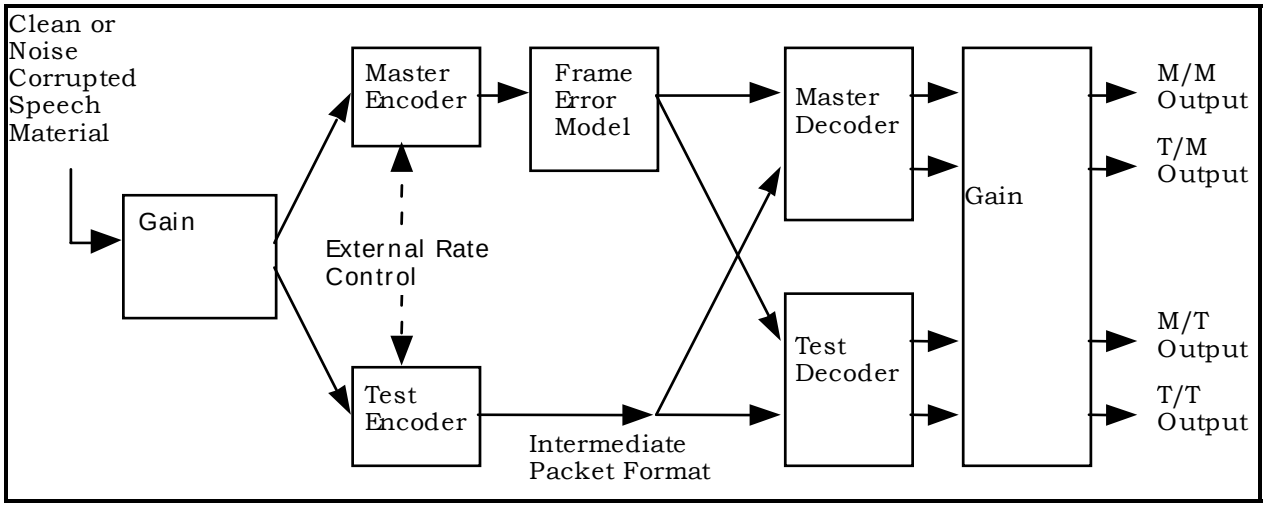


Figure 1.3-1 Test Model

Various implementations of the encoder and decoder, especially those in hardware, may not be designed to deliver or accept a continuous data stream as previously described. It is the responsibility of the manufacturer to implement a test platform that is capable of delivering and accepting these formats in order to complete the performance tests described in the following sections. This may involve a custom hardware interface or a fair implementation of the algorithm in software, or some other mechanism. A fair implementation in software shall yield bit-exact output with reference to any hardware implementation that it is claimed to represent.

The input speech material has been precision-limited by an 8-bit μ-law quantization algorithm in which the inverse quantized linear samples fill the entire 16-bit linear range. As specified within Section 3 of [1], the master codec assumes a 16-bit integer input/output normalization.

2 CODEC MINIMUM STANDARDS[2]

This section describes the validation procedures that shall be used to verify the quality and interoperability of an EVRC implementation. The procedures are both comprehensive and backward compatible, in that they are provided for the SO 3, SO 68, SO70, and SO 73 implementations of EVRC. The validation procedures comprise a set of objective and subjective tests as well as a maximum algorithmic delay Recommendation. These are described in the following sections:

2.1 Performance Testing for SO 3

2.1.1 Objective Performance Testing for SO 3

The objective testing portion of this specification consists of an average data rate test, and compliance to End-to-End Algorithmic Delay and Unity-gain requirements.

2.1.1.1 Average Data Rate Test

The average data rate for the test codec shall be measured using benchmark files that are contained on the accompanying Software Distribution (in the /so3/objctv subdirectory).

The average data rate for the test codec shall be measured using twelve benchmark files that are contained in the associated Software Distribution (in the /so3/objctv subdirectory). Each file exhibits a different combination of input level: -12 dB, -22 dB, and -32 dB, and background noise conditions: ambient background noise, 20 dB SNR babble noise condition, 15 dB SNR car noise condition and 12 dB SNR street noise. The background noise has been introduced by mixing the clean speech recording with the noise recording at the appropriate levels. The benchmark recording employed in the average data rate test is a single-sided recording similar to a telephone conversation. It exhibits an approximate voice activity factor of 0.35. The processed files are not used in the subjective portion of the experiment. The length of each of the benchmark files is approximately 480 seconds.

2.1.1.1.1 Average Data Rate Computation

The average data rate for the test codec shall be computed for each of the benchmark files as follows:

$$R = (9600*N_1 + 4800*N_2 + 1200*N_8)/N,$$

where

N_1 = number of frames encoded at Rate 1,

N_2 = number of frames encoded at Rate 1/2,

N_8 = number of frames encoded at Rate 1/8, and

$N = N_1 + N_2 + N_8$.

The total average data rate for the test codec is then given by:

[2] This section does not apply whenever a codec has demonstrated bit-exactness. See 3.1.4, 3.2.4, 3.3.4, or 3.4.4.

R_{avg} = .0833 * {R(babble noise segment @ -12dB) + R(car noise segment @ -12dB).+ R(street noise segment @ -12dB) + R(ambient background segment @ -12dB).+ R(babble noise segment @ -22dB) + R(car noise segment @ -22dB).+
R(street noise segment @ -22dB) + R(ambient background segment @ -22dB).+
R(babble noise segment @ -32dB) + R(car noise segment @ -32dB).+
R(street noise segment @ -32dB) + R(ambient background segment @ -32dB)}.

See Section 3.1.2.1 for details in using the provided software tool that can be used to aid in making this calculation.

2.1.1.1.2 Average Data Rate Requirement

The total average data rate R_{avg} shall not exceed 4400 bps, otherwise the test codec fails the compliance test.

2.1.1.2 Unity Gain Requirement

The specific EVRC test codec shall output speech with unity gain when compared with the input speech. The unity gain measurement (output active–speech level/input active speech level) will be performed over the entire input speech database for the clean, nominal-level source conditions for each mode. The measurement should be made using the STL-2000 tool [6] [6a] actlev, and must not show more than ± 0.5 dB deviation between input and output active speech levels. This procedure is fully described in [9].

2.1.1.3 End-to-end Algorithmic Delay Recommendation

The algorithmic delay for the specific EVRC test codec should be calculated analytically by the codec manufacturer. In considering the algorithmic delay, it can be assumed that all transmission channels have infinite bandwidth, and that all processing elements have infinite throughput. Algorithmic delay is defined as the sum of all sequential filter delays and buffering delays in the encode/decode path.

The maximum end-to-end algorithmic delay should be no greater than that of the master codec. For the master codecs defined in [1], the algorithmic delay is given as:

Delay Element	SO 3
Signal Preprocessing Delay:	3 milliseconds
LPC Analysis “Look-ahead”:	10 milliseconds
LPC Analysis Window:	20 milliseconds
Total:	33 milliseconds

Therefore, the total algorithmic delay imposed by a SO 3 test codec should not exceed 33 milliseconds.

2.1.2 Subjective Performance Testing for SO 3

This section outlines the subjective testing methodology of the subjective performance test. The purpose of this testing is to evaluate the quality of the test codec under a variety of conditions which may occur in the CDMA system. To accomplish this, two listening experiments have been designed to test speech codec quality under a variety of conditions. These conditions include channel impairments, codec tandem, audio background noise, and different input levels. In addition, half-rate maximum operation of the codec will be examined.

2.1.2.1 Definition

The codec subjective test is intended to validate the implementation of the speech codec being tested using the master codec defined in Section 3.1.3 as a reference. The subjective tests for SO 3 are based on the Absolute Category Rating, Mean Opinion Score (MOS) test as described in [10].

2.1.2.2 Method of Measurement

The subjective test involves a listening-only assessment of the quality of the codec being tested, using the master codec as a reference. Subjects from the general population of telephone users will rate the various conditions of the test. Material supplied with this standard for use with this test includes source speech, impaired packet files from the master codec encoder, and source speech processed by various Modulated Noise Reference Unit (MNRU) conditions and other references. The basic Absolute Category Rating test procedure involves rating all conditions using a five-point scale describing the opinion of the test condition. This procedure is fully described in [10].

2.1.2.3 Test Conditions and Test Design for SO 3 Listening Experiments

The two listening experiments for SO 3 are similar in design, and are performed as MOS listening tests. Each experiment will test the same number of codecs, and the number of test conditions for each experiment is five. There will be one condition typifying CDMA channels (3% FER), a clear channel condition and a clear channel tandem condition. All tandem conditions shall be asynchronous, where asynchronous implies the introduction of a partial frame offset between encoding operations. A nominal input level of -22 dB shall be used for these conditions. Additional test conditions include background noise and audio input level variation.

For reference, μ-law, 4 MNRU conditions (5, 15, 20 and 25 dBQ values) and G.728 [8] will be included in each experiment. The IS-96-C codec, specified in [2], is included for all conditions as an additional codec.

2.1.2.3.1 Subjective Experiment I for SO 3

The Test Conditions for Listening Experiment I are presented in Table 2.1.2.3.1-1.

Table 2.1.2.3.1-1 SO 3 Listening Experiment I Conditions

Condition	Description
Type of test	MOS (P.800)
Number of talkers	4 males, 4 females
Background noise	none (ambient)
Audio Input level	-22 dB (except for high/low input cond)
Filter characteristics	IRS
Reference conditions	μ-law source, 5, 15, 20, 25 dBQ, G.728

Condition	Description
Test conditions	(1) Clean (2) High Audio Input Level -12 dB (3) Low Audio Input Level -32 dB (4) 3% FER (forward and reverse) (5) Rate 1/2 Maximum
Number of codecs	(5) M/M, T/M, M/T, T/T, IS-96-C
Encoding stages	single

The Test Design for Listening Experiment I are presented in Table 2.1.2.3.1-2.

Table 2.1.2.3.1-2 SO 3 Listening Experiment I Design

Label	Operating Point	Condition	Enc/Dec Connection
a01	EVRC-A	Clean, Nominal, -22 dB	M-M
a02	EVRC-A	Clean, Nominal, -22 dB	M-T
a03	EVRC-A	Clean, Nominal, -22 dB	T-M
a04	EVRC-A	Clean, Nominal, -22 dB	T-T
a05	IS-96-C	Clean, Nominal, -22 dB	R-R
a06	EVRC-A	High, -12 dB	M-M
a07	EVRC-A	High, -12 dB	M-T
a08	EVRC-A	High, -12 dB	T-M
a09	EVRC-A	High, -12 dB	T-T
a10	IS-96-C	High, -12 dB	R-R
a11	EVRC-A	Low, -32 dB	M-M
a12	EVRC-A	Low, -32 dB	M-T
a13	EVRC-A	Low, -32 dB	T-M
a14	EVRC-A	Low, -32 dB	T-T
a15	IS-96-C	Low, -32 dB	R-R
a16	EVRC-A	3% FER For & Rev	M-M
a17	EVRC-A	3% FER For & Rev	M-T
a18	EVRC-A	3% FER For & Rev	T-M
a19	EVRC-A	3% FER For & Rev	T-T
a20	IS-96-C	3% FER For & Rev	R-R
a21	EVRC-A, HR-Max	Nominal, -22 dB	M-M
a22	EVRC-A, HR-Max	Nominal, -22 dB	M-T
a23	EVRC-A, HR-Max	Nominal, -22 dB	T-M
a24	EVRC-A, HR-Max	Nominal, -22 dB	T-T

Label	Operating Point	Condition	Enc/Dec Connection
a25	IS-96-C, HR-Max	Nominal, -22 dB	R-R
a26	Reference	MNRU 5dB	
a27	Reference	MNRU 15dB	
a28	Reference	MNRU 20dB	
a29	Reference	MNRU 25dB	
a30	Reference	G.728	
a31	Reference	u-Law Source	

2.1.2.3.2 Subjective Experiment II for SO 3

The Test Conditions for Listening Experiment II are presented in Table 2.1.2.3.2-1.

Table 2.1.2.3.2-1 SO 3 Listening Experiment II Conditions

Condition	Description
Type of test	MOS (P.800)
Number of talkers	4 males, 4 females
Background noise	ambient and specified test conditions
Audio Input level	-22 dB
Filter characteristics	flat voice
Reference conditions	µ-law source, 5, 15, 20, 25 dBQ , G.728
Test conditions	(1) Clean (2) Car Noise (IRS) at 15 dB S/N (3) Street Noise (flat) at 12 dB S/N (4) Office Babble (flat) at 20 dB S/N (5) Tandem
Number of codecs	(5) M/M, T/M, M/T, T/T, IS-96-C
Encoding stages	single and tandem

The Test Design for Listening Experiment II are presented in Table 2.1.2.3.2-2.

Table 2.1.2.3.2-2 SO 3 Listening Experiment II Design

Label	Operating Point	Condition	Enc/Dec Connection
b01	EVRC-A	Clean, Nominal, -22 dB	M-M
b02	EVRC-A	Clean, Nominal, -22 dB	M-T
b03	EVRC-A	Clean, Nominal, -22 dB	T-M
b04	EVRC-A	Clean, Nominal, -22 dB	T-T

Label	Operating Point	Condition	Enc/Dec Connection
b05	IS-96-C	Clean, Nominal, -22 dB	R-R
b06	EVRC-A	Car Noise (IRS) at 15 dB S/N	M-M
b07	EVRC-A	Car Noise (IRS) at 15 dB S/N	M-T
b08	EVRC-A	Car Noise (IRS) at 15 dB S/N	T-M
b09	EVRC-A	Car Noise (IRS) at 15 dB S/N	T-T
b10	IS-96-C	Car Noise (IRS) at 15 dB S/N	R-R
b11	EVRC-A	Street Noise (Flat) at 12 dB S/N	M-M
b12	EVRC-A	Street Noise (Flat) at 12 dB S/N Car Noise (IRS) at 12 dB S/N	M-T
b13	EVRC-A	Street Noise (Flat) at 12 dB S/N Car Noise (IRS) at 12 dB S/N	T-M
b14	EVRC-A	Street Noise (Flat) at 12 dB S/N Car Noise (IRS) at 12 dB S/N	T-T
b15	IS-96-C	Street Noise (Flat) at 12 dB S/N Car Noise (IRS) at 12 dB S/N	R-R
b16	EVRC-A	Office Noise (Flat) at 20 dB S/N Car Noise (IRS) at 15 dB S/N	M-M
b17	EVRC-A	Office Noise (Flat) at 20 dB S/N Car Noise (IRS) at 15 dB S/N	M-T
b18	EVRC-A	Office Noise (Flat) at 20 dB S/N Car Noise (IRS) at 15 dB S/N	T-M
b19	EVRC-A	Office Noise (Flat) at 20 dB S/N Car Noise (IRS) at 15 dB S/N	T-T
b20	IS-96-C	Office Noise (Flat) at 20 dB S/N Car Noise (IRS) at 15 dB S/N	R-R
b21	EVRC-A, Tandem	Nominal, -22 dB	M-M,/M-M
b22	EVRC-A, Tandem	Nominal, -22 dB	M-M/T-T
b23	EVRC-A, Tandem	Nominal, -22 dB	T-T/M-M
b24	EVRC-A, Tandem	Nominal, -22 dB	T-T/T-T
b25	IS-96-C, Tandem	Nominal, -22 dB	R-R
b26	Reference	MNRU 5dB	
b27	Reference	MNRU 15dB	
b28	Reference	MNRU 20dB	
b29	Reference	MNRU 25dB	
b30	Reference	G.728	
b31	Reference	u-Law Source	

1

2.1.2.3.3 Numerical Parameters for SO 3 Listening Experiments

Table 2.1.2.3.3-1 describes the resultant numerology that is used for each of the two SO 3 listening experiments. The first column is a variable name given to each of the parameters, the second column is the description of the parameter, the third column shows the required calculation for determining the value of the parameter if it is dependent upon other parameter values and the last column shows the numerical value for each of the parameters. For each listening experiment, four codecs plus the IS-96-C codec are evaluated. The number of reference conditions in each of the two listening experiments is six, and the number of test conditions is five.

Table 2.1.2.3.3-1 Numerical Parameters for SO 3 Listening Experiments

Var	Parameter	Calculation	Experiment I Value	Experiment II Value
C1	Codecs		5	5
C2	Codec Test Conditions		5	5
C3	Reference Conditions		6	6
C4	Total Conditions	C1*C2+C3	31	31
C5	Talkers		8	8
C6	Stimuli per Talker		8	8
C7	Stimuli per Condition	C5*C6	64	64
C8	Total Stimuli per Experiment	C4*C7	1984	1984
C9	File Sessions		8	8
C10	Stimuli per File Session	C8/C9	248	248
C11	Listeners (Voters)		64	64
C12	Listeners (Voters) per File Session	C11/C9	8	8
C13	Votes per Condition	C9*C10*C12/C4	512	512

2.1.3 Source Speech Material for SO 3 Testing

All source material is derived from the Harvard Sentence Pair Database and matched in overall level.

There are a total of 64 original source files from 8 different talkers. While individual sentences are repeated, every sample uses a distinct sentence pairing. Talkers were chosen to have distinct voice qualities and are native speakers of North American English.

For the following discussion, it may be useful to refer to Table 4-1 for the configuration of the associated Software Distribution.

2.1.3.1 Source Speech Material for Experiment I

The source speech material for subjective Experiment I is contained in directory /so3/subjctv/exp1/source. Each sentence is IRS filtered, gain adjusted, and μ-Law companded in

accordance with [7]. The talkers in subjective Experiment I consist of four adult males and four adult females.

The source material for Experiment I consists of 8 sentence pairs from 8 different speakers for a total of 64 speech files for both of the nominal input conditions (conditions 1 and 5). These files are named *.s22. This directory also contains the source material for each of the high and low level input conditions, which are named *.s12 and *.s32, respectively, for a total of 3 x 64 = 192 files. The speech database also includes samples processed through the various reference conditions in directory /so3/subjctv/exp1/ref. The reference conditions are named *.q05 through *.q25 for the respective MNRU conditions and *.728 for the G.728 reference. The samples processed by the IS-96-C codec for each of the five conditions are named *.qc1 through *.qc5, respectively and *.qc4 is replaced with *.qf3 and *.qr3 corresponding to the IS-96-C codec 3% forward and reverse FER, respectively, also reside here.

2.1.3.2 Source Speech Material for Experiment II

The source speech material for subjective Experiment II is contained in directory /so3/subjctv/exp2/source. Each sentence is flat filtered and μ-law companded in accordance with [7]. The talkers in subjective Experiment II consist of four adult males and four adult females.

The clean source material for Experiment II, conditions 1 and 5, consists of 8 sentence pairs from 8 different speakers for a total of 64 speech files. These files are named *.s22. This directory also contains the source material for the car, street, and babble noise conditions, which are named *.car, *.str, and *.bab, respectively, for a total of 4 x 64 = 256 files. The speech database also includes samples processed through the various reference conditions in directory /so3/subjctv/exp2/ref. The reference conditions are named *.q05 through *.q25 for the respective MNRU conditions and *.728 for the G.728 reference. The samples processed by the IS-96-C codec for each of the five conditions (named *.qc1 through *.qc5, respectively) also reside here.

2.1.4 Processing of Speech Material for SO 3 Testing

The source speech material shall be processed by the various combinations of encoders and decoders listed in the descriptions of the two experiments given in Section 2.1.2. The master codec software described in Section 3.1.3 shall be used in the processing involving the master codec. Generally, the master codec encoder and decoder outputs have been provided in the respective /so3/subjctv/exp*/m_pkt and /so3/subjctv/exp*/m_m directories. Execution of the master codec software is generally needed only for the test encoder/master decoder combination for each experiment/condition. The exception to this is the tandem condition in Experiment II, where double codec processing is required (see Section 2.1.4.4).

All codec processing shall be done digitally. Noise suppression and post-filter options shall be enabled for both the master and the test codecs. The digital format of the speech files is described in Section 3.1.4.4.

The naming convention of the processed speech is as follows: For the packet files in the /so3/subjctv/exp1/m_pkt directory (Experiment I), the *.p12 files are the master packet files for the *.s12 source files. Likewise, the *.p22 and *.p32 files are the respective packet files for the *.s22 and *.s32 source files. The *.pf3 and *.pr3 are the impaired packet files which will be described in Section 2.1.4.3. Condition five (Rate 1/2 maximum), it uses *.phr as the extension for the half rate max packets.

Similarly, the directory /so3/subjctv/exp2/m_pkt contains the master packet files for Experiment II. Here, the *.p22 files are the master packet files for the *.s22 source files, and the *.pc, *.pb, and *.ps files are the master packet files for the *.car, *.bab, and *.str source files, respectively.

For the master encode/master decode directories (/so3/subjctv/exp*/m_m), the naming convention of the speech files is such that the first two characters of the suffix indicate the codec combination and third indicates the condition number (1 through 5). It is required that this convention be used for the other codec combinations (mt, tm, and tt) so that the supplied randomization lists (see Section 2.1.5) are valid. Two exceptions to this naming convention is the master encoder/master decoded 3% reverse link FER files which shall be assigned the extension *.tm4 and the 3% forward link FER files shall be assigned the extension *.mm4.

2.1.4.1 Encoding by the Test Codec

All of the source files will be encoded by the test codec to produce encoded packet files. For ease of reference, it is recommended that directories /so3/subjctv/exp1/t_pkt and /so3/subjctv/exp2/t_pkt be created to deposit the test encoder output packets, and that the naming conventions be made consistent with the master codec.

2.1.4.2 Decoding by the Master/Test Codecs

The encoded packet files generated from the various encoders/conditions shall be processed through the master and test decoders. For all conditions, the signal power shall be normalized to -22 dB. The signal shall then be μ-law companded into PCM files. See Sections 3.1.2.2 and 3.1.2.3 for details in using the provided software tools that can be used for this post-processing.

2.1.4.3 Introduction of Impairments

For the 3% frame error condition (Experiment I, condition 4), the impaired master codec encoded packet files are provided in the /so3/subjctv/exp1/m_pkt directory. Unlike other conditions, this condition uses only the test decoder and not the test encoder. The performance of the test decoder is compared to that of master decoder using master encoder generated packets from two different frame error models: 3% forward FER and 3% reverse FER. The 3% forward FER packets (*.pf3) are then used by the test decoder to generate the master encoder/test decoder combination (*.mt4), and the 3% reverse FER packets (*.pr3) are used by the test decoder to generate the master encoder/test decoder combination (*.tt4). The respective master decoder outputs are the *.mm4 and *.tm4.

To clarify the naming convention, the following four conditions are tested:

- ***.mm4 -** master encoder, master decoder, 3% forward link FER
- ***.tm4 -** master encoder, master decoder, 3% reverse link FER
- ***.mt4 -** master encoder, test decoder, 3% forward link FER
- ***.tt4 -** master encoder, test decoder, 3% reverse link FER

2.1.4.4 Tandem Conditions

The clear channel tandem condition shall be performed by:

- encoding the appropriate source file,
- decoding the encoder's output file,

- normalizing signal power to -22dB,
- companding the modified decoded speech file to μ-law PCM format,
- encoding the μ-law PCM companded version of the decoded speech file,
- decoding the resultant encoder's output file to generate the processed speech file,
- normalizing signal power to -22dB,
- companding the modified decoded speech file to μ-law PCM.

This process is performed for each combination of master encode/test decode, test encode/master decode, and test encode/test decode. The master/test combinations for tandem processing represent master encode/test decode/master encode/test decode and vice versa for the test/master combination. The master encode/master decode files are provided.

The following four conditions are tested:

- M/M + M/M
- M/T + M/T
- T/M + T/M
- T/T + T/T

To expedite processing, it may be possible to use the output files for Experiment II condition 1 (*.tm1, *.mt1, and *.tt1) as the input for the three test combinations.

It is also worth noting that the front-end algorithmic delay through the master codec is 13 ms (or 104 samples), which can be accounted for by the noise suppression overlap delay plus the LPC look-ahead. This 13 ms delay will ensure the proper tandem processing. It may be beneficial for the test codec to incur the same delay as the master codec to avoid potential quality differences due to framing skew. This kind of delay ensures asynchronous tandem processing.

2.1.4.5 Rate 1/2 Maximum Processing

The appropriate speech files will be processed through the codecs for the Rate 1/2 Maximum processing test conditions. The test speech codec shall be constrained to operate such that Rate 1 coding is not used.

2.1.4.6 Ensuring Proper Encoded Frame Packet Files

All encoded frame packet files shall be examined to ensure that the files only contain data in those file locations where data should exist for a given data rate.

The examination of the encoded frame packet files should indicate the occurrence of any improper data in the files but the examination must not alter the encoded frame packet files in any way.

2.1.5 Randomization

For each of the two subjective experiments, each presentation sample consists of one sentence pair processed under a condition of the test. The samples shall be presented to the listeners in a random order. The listeners for each file set shall be presented with practice trials for subjective Experiments

I and II. The randomization of the test samples has been constrained in the following ways for the two experiments:

1. A test sample for each codec combination, talker and level, channel condition, or background noise level (Experiment I or II) or MNRU value and talker shall be presented exactly once.
2. Randomization has been done in “blocks”, such that one sample of each codec/level, codec/channel condition, or codec/background noise level (again depending on Experiment I or II) or MNRU value will be presented once, with a randomly selected talker, in each block. This ensures that listeners rate each codec/condition being tested equally often in the initial, middle and final parts of the session and will mitigate the effects of practice and fatigue. A block contains 31 file samples. A “session” will consist of eight blocks of 31 file samples (plus one practice block of 31 at the beginning of each session) for each experiment. There are a total of eight sessions per experiment. A particular randomization session shall not be presented to more than eight listeners.
3. Talkers shall be chosen so that the same talker is never presented on two consecutive trials within the same block.

The randomization lists for each of the eight file sets of each experiment are given in /so3/subjctv/exp1/data/play*.lst and /so3/subjctv/exp2/data/play*.lst, respectively.

2.1.6 Presentation

Presentation of speech material for the SO 3 codec listening tests shall be made with one side of high fidelity circum-aural headphones. The speech material delivery system shall meet the requirements of Section 3.1.1.1. The delivery system shall be calibrated to deliver an average listening level of -16 dBPa (78 dB SPL). The equivalent acoustic noise level of the delivery system should not exceed 35 dBA as measured on a standard A-weighted meter.

The listeners should be seated in a quiet room, with an ambient noise of 40 dBA or below.

2.1.7 Listeners

The listener sample is intended to represent the population of telephone users with normal hearing acuity. The listeners should be naïve with respect to telephony technology issues; that is, they should not be experts in telephone design, digital voice encoding algorithms, and so on. They should not be trained listeners; that is, they should not have been trained in these or previous listening studies using feedback trials. The listeners should be adults of mixed sex and age.

Each listener shall provide data only once for a particular evaluation. A listener may participate in different evaluations, but test sessions performed with the same listener should be at least one month apart so as to reduce the effect of cumulative experience.

2.1.8 Listening Test Procedures

The listeners shall listen to each sample and rate the quality of the test sample using a five-point scale, with the points labeled:

1. Bad
2. Poor

3. Fair
4. Good
5. Excellent

Data from 64 listeners shall be used for each of the two experiments. The experiment may be run with up to eight listeners in parallel; that is, hearing the same random order of test conditions at the same time.

Before starting the test, the listeners should be given the instructions in Figure 2.1.8-1. The instructions may be modified to allow for variations in laboratory data-gathering apparatus.

This is an experiment to determine the perceived quality of speech over the telephone. You will be listening to a number of recorded speech samples, spoken by several different talkers, and you will be rating how good you think they sound.

The sound will appear on one side of the headphones. Use the live side on the ear you normally use for the telephone.

On each trial, a sample will be played. After you have listened to each passage, the five buttons on your response box will light up. Press the button corresponding to your rating for how good or bad that particular passage sounded.

During the session you will hear samples varying in different aspects of quality. Please take into account your total impression of each sample, rather than concentrating on any particular aspect.

The quality of the speech should be rated according to the scale below:

Bad	Poor	Fair	Good	Excellent
1	2	3	4	5

Rate each passage by choosing the word from the scale which best describes the quality of speech you heard. There will be 279 trials, including 31 practice trials at the beginning.

Thank you for participating in this research.

Figure 2.1.8-1 Instructions for Listeners

2.1.9 Analysis of Results

The response data from the practice blocks shall be discarded. Data sets with missing responses from listeners shall not be used. Responses from the different sets of encoder/decoder processed files shall be treated as equivalent in the analysis.

The votes for each of the 31 conditions and references for each of SO 3 Experiment I and II shall be averaged in accordance with [10], to produce an associated mean opinion score (MOS). Additionally, the standard error (SER) for each condition shall be calculated as described in the next section.

2.1.10 Minimum Subjective Requirement

For each of the test combinations (T/M, M/T, T/T), the MOS results are compared to those of the respective master codec (M/M). (The exception to this being the 3% FER case in which M/T is compared to M/M and T/T is compared to T/M[3])

If the MOS for the test combination/condition is within an allowable difference (as defined below) of the MOS for the master combination/condition, then the subjective test is passed for that combination/condition. If any of the test combinations/conditions exceeds the maximum allowable difference, the test codec fails the compliance test.

These requirements can be clarified by first defining the MOS for a given combination/condition as:

$$MOS(i,j,k) = \frac{1}{512}\sum_{n=1}^{512} v(i,j,k,n), \quad \begin{cases} i \in \{1,2\} \\ j \in \{1,...,5\} \\ k \in \{1,...,4\} \end{cases}$$

$$MOS(i,j,k) = \frac{1}{512}\sum_{n=1}^{512} v(i,j,k,n), \quad \begin{cases} i \in \{1,2\} \\ j \in \{1,...,5\} \\ k \in \{1,...,4\} \end{cases} \qquad \text{(2.1.10-1)}$$

where *i* is the experiment number, *j* is the condition number, *k* is the codec combination number (1 = M/M with 3% forward link FER, 2 = M/T with 3% forward link FER, 3 = T/M with 3% reverse link FER, 4 = T/T with 3% reverse link FER), and *v* is the associated listener vote.

Then, the per combination/condition requirement can be defined as:

[3] Refer to Section 2.1.4.3. In this case, M/M and M/T are, respectively, the outputs of the master and test decoders in response to packets generated by the master encoder that have been corrupted using a 3% forward link FER model. Similarly, T/M and T/T are the outputs of the master and test decoders in response to packets generated by the master encoder that have been corrupted using a 3% reverse link error model.

$$MOS(i,j,1) - MOS(i,j,k') \le \delta(i,j,k'), \quad \begin{cases} i \in \{1,2\} \\ j \in \{1,...,5\} \\ k' \in \{2,...,4\} \\ \{i,j\} \neq \{1,4\} \end{cases}$$

$$MOS(i,j,1) - MOS(i,j,k') \le \delta(i,j,k'), \quad \begin{cases} i \in \{1,2\} \\ j \in \{1,...,5\} \\ k' \in \{2,...,4\} \\ \{i,j\} \neq \{1,4\} \end{cases} \qquad \text{(2.1.10-2)}$$

except for the 3% FER condition (i = 1, j = 4) where the following requirement is defined:

$$MOS(1,4,k'') - MOS(1,4,k''+1) \le \delta(1,4,k''+1), \quad k'' \in \{1,3\}$$
$$MOS(1,4,k'') - MOS(1,4,k''+1) \le \delta(1,4,k''+1), \quad k'' \in \{1,3\}. \qquad \text{(2.1.10-3)}$$

In Equation 2.1.10-2, the maximum allowable difference $\delta(i, j,k')$ is given by:

$$\delta(i,j,k') = \max\left(0.12,\ c(i,j,k')\sqrt{SER^2(i,j,1) + SER^2(i,j,k')}\right)$$
$$\delta(i,j,k') = \max\left(0.12,\ c(i,j,k')\sqrt{SER^2(i,j,1) + SER^2(i,j,k')}\right) \qquad \text{(2.1.10-4)}$$

Similarly, in Equation 2.1.10-3, the maximum allowable difference $\delta(i, j,k''+1)$is given by:

$$\delta(i,j,k''+1) = \max\left(0.12, c(i,j,k''+1)\sqrt{SER^2(i,j,k'') + SER^2(i,j,k''+1)}\right)$$
$$\delta(i,j,k''+1) = \max\left(0.12, c(i,j,k''+1)\sqrt{SER^2(i,j,k'') + SER^2(i,j,k''+1)}\right) \qquad \text{(2.1.10-5)}$$

where i, j,k' and k'' are as defined above and the multipliers $c(i,j,k)$ are given in Table 2.1.10-1. The standard errors $SER(i,j,k)$ for each condition are defined as:

$$SER(i,j,k) = \sqrt{\frac{\sum_{n=1}^{512}\left(v(i,j,k,n) - MOS(i,j,k)\right)^2}{261632}}$$
$$SER(i,j,k) = \sqrt{\frac{\sum_{n=1}^{512}\left(v(i,j,k,n) - MOS(i,j,k)\right)^2}{261632}} \qquad \text{(2.1.10-6)}$$

Specifically stating the requirement, Equations 2.1.10-2 and 2.1.10-3 shall be true for all cases; otherwise the test codec fails the compliance test.

Table 2.1.10-1 Multipliers for Equations 2.1.10-4 and 2.1.10-5

Experiment	Condition	Description	$c(i, j, k)$		
i	*j*		M/T (*k*=2)	T/M (*k*=3)	T/T (*k*=4)
I	1	Clean	2.64	2.73	3.04
	2	High Audio Input Level	3.09	3.67	4.08
	3	Low Audio Input Level	2.60	2.94	3.38
	4	3% FER	3.96	N/A	3.34
	5	Rate 1/2 Maximum	2.58	2.65	2.70
II	1	Clean	2.00	3.17	3.65
	2	Car Noise	2.00	2.00	2.00
	3	Street Noise	2.00	2.00	2.00
	4	Office Babble	2.26	3.22	3.49
	5	Tandem	2.56	3.88	4.65

2.1.11 Expected Results for Reference Conditions

The MNRU conditions have been included to provide a frame of reference for the MOS test. Also, they provide anchor conditions for comparing results between test laboratories. In listening evaluations where test conditions span approximately the same range of quality, the MOS results for similar conditions should be approximately the same. Data from previous studies allows a generalization to be made concerning the expected MOS results for the MNRU reference conditions (see Figure 2.1.11-1).

MOS scores obtained for the MNRU conditions in any SO 3 validation test should be compared to those shown in the graph below. Inconsistencies beyond a small shift in the means in either direction or a slight stretching or compression of the scale near the extremes may imply a problem in the execution of the evaluation test. In particular, MOS should be monatomic with MNRU, within the limits of statistical resolution; and the contour of the relation should show a similar slope.

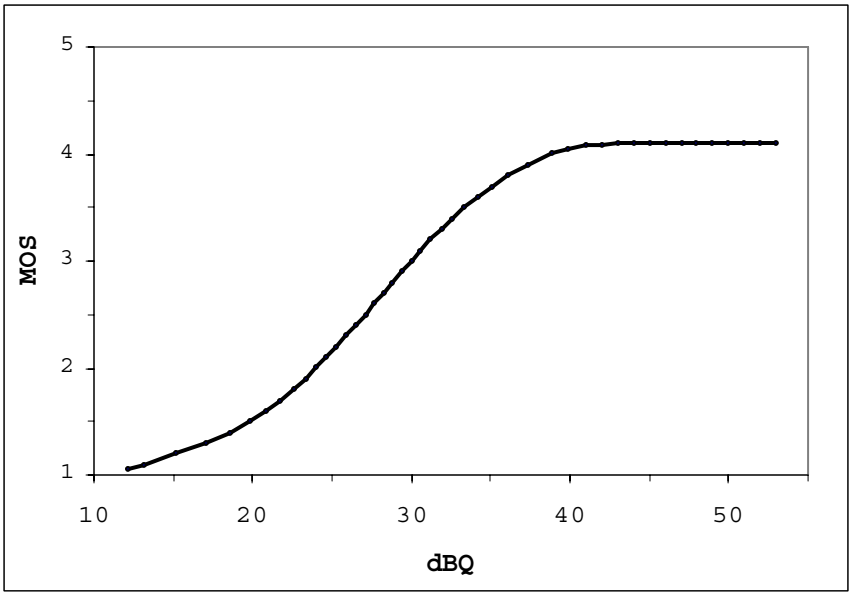


Figure 2.1.11-1 MOS versus MNRU

2.2 Performance Testing for SO 68

2.2.1 Objective Performance Testing for SO 68

The objective testing portion of this specification consists of an average data rate test, and compliance to End-to-End Algorithmic Delay and Unity-gain requirements.

2.2.1.1 Average Data Rate Test

The average data rate for the test codec shall be measured using six source speech files that are contained in the /so68/subjctv/exp*/source/ directory. Each file exhibits a different condition: power levels: -12 dB, -22 dB, and -32 dB, and background noise conditions: 20 dB SNR babble noise condition, 15 dB SNR car noise condition and 15 dB SNR street noise. The input source files used in the average data rate test have an approximate voice activity factor of 0.78, and are the same input files used in the subjective portion of the experiment.

2.2.1.1.1 Average Data Rate Computation for SO 68

The average channel data rate for the test codec shall be computed for each of the benchmark files as follows:

$$R = (9600*N_1 + 4800*N_2 + 2400*N_4 + 1200*N_8)/N,$$

where

N_1 = number of frames encoded at Rate 1,

N_2 = number of frames encoded at Rate 1/2,

N_4 = number of frames encoded at Rate 1/4,

N_8 = number of frames encoded at Rate 1/8, and

$N = N_1 + N_2 + N_4 + N_8$.

The total average channel data rate for the test codec is then given by:

R_{avg} = 1/6 * { R(ambient background segment @ -12dB).+ R(ambient background segment @ -32dB).+ R(ambient background segment @ -22dB). + R(20 dB SNR babble noise segment @ -22dB) + R(15 dB SNR car noise segment @ -22dB).+ R(15 dB SNR street noise segment @ -22dB) }.

The above files are to be processed with EVRC-B encoder at various capacity operating points (defined by the active speech average channel rate) shown in Table 2.2.1.1.1-1.

Table 2.2.1.1.1-1 Target ADR vs Capacity Operating Point

Capacity Operating Point (active speech average channel data rate)	Target Average Channel Data Rate, kbps
EVRC-B (9.3k bits/sec)	6.93 (+1.5%)
EVRC-B (8.5 bits/sec)	6.42 (+1.5%)
EVRC-B (7.5k bits/sec)	5.52 (+1.5%)

Capacity Operating Point (active speech average channel data rate)	Target Average Channel Data Rate, kbps
EVRC-B (7.0k bits/sec)	5.24 (+1.5%)
EVRC-B (6.6k bits/sec)	4.82 (+1.5%)
EVRC-B (6.2k bits/sec)	4.62 (+1.5%)
EVRC-B (5.8k bits/sec)	4.45 (+1.5%)
EVRC-B (Half-Rate Max, 4.8k bits/sec)	3.75 (+1.5%)

The above table provides the maximum allowable average channel rate (including full, half, quarter, and eighth-rate) for the different capacity operating points. These maximum allowable average channel rates were obtained by processing the 6 bench mark files through the master floating point software. See Section 3.2.2.1 for details in using the provided software tool that can be used to aid in making this calculation.

2.2.1.1.2 Average Data Rate Requirement for SO 68

The total average data rate R_{avg} for each capacity operating point shall not exceed the target average data rate by more than the tolerance level in Table 2.2.1.1.1-1, otherwise the test codec fails the compliance test.

2.2.1.2 Unity Gain Requirement

The specific EVRC-B test codec shall output speech with unity gain when compared with the input speech. The unity gain measurement (output active–speech level/input active speech level) will be performed over the entire input speech database for the clean, nominal-level source conditions for each mode. The measurement should be made using the STL-2000 tool [6] [6a] actlev, and must not show more than $\pm$ 0.5 dB deviation between input and output active speech levels. This procedure is fully described in [9].

2.2.1.3 End-to-end Algorithmic Delay Recommendation

The algorithmic delay for the specific EVRC-B test codec should be calculated analytically by the codec manufacturer. In considering the algorithmic delay, it can be assumed that all transmission channels have infinite bandwidth, and that all processing elements have infinite throughput. Algorithmic delay is defined as the sum of all sequential filter delays and buffering delays in the encode/decode path.

The maximum end-to-end algorithmic delay should be no greater than that of the master codec. For the master codecs defined in [1], the algorithmic delay is given as:

Delay Element	SO 68
Signal Preprocessing Delay:	3 milliseconds
LPC Analysis “Look-ahead”:	10 milliseconds
LPC Analysis Window:	20 milliseconds
Total:	33 milliseconds

Therefore, the total algorithmic delay imposed by a SO 68 test codec should not exceed 33 milliseconds.

2.2.2 Subjective Performance Testing for SO 68

This section outlines the subjective testing methodology of the subjective performance test. The purpose of this testing is to evaluate the quality of the test codec under a variety of conditions which may occur in the CDMA system. To accomplish this, two listening experiments have been designed to test speech codec quality under a variety of conditions. These conditions include channel impairments, audio background noise, and different input levels

2.2.2.1 Definition

The codec subjective test is intended to validate the implementation of the speech codec being tested using the master codec defined in 3.2.3 as a reference. Experiment I is based on the Absolute Category Rating (ACR) method, which yields the Mean Opinion Score (MOS) as described in [10]. Experiment II is based on the ITU-T Recommendation P.835 described in [13].

2.2.2.2 Method of Measurement

The subjective test involves a listening-only assessment of the quality of the codec being tested, using the master codec as a reference. Subjects from the general population of telephone users will rate the various conditions of the test. Material supplied with this standard for use with this test includes source speech, impaired packet files from the master codec encoder, and source speech processed by various Modulated Noise Reference Unit (MNRU) conditions and other references. The basic Absolute Category Rating test procedure involves rating all conditions using a five-point scale describing the opinion of the test condition. This procedure is fully described in [10]. The P.835 test method involves rating all conditions on scales of “Signal”, “Background”, and “Overall” quality and is fully described in [13].

2.2.2.3 Test Conditions and Test Design for SO 68

The first listening experiment for SO 68 is performed as an ACR listening test. The second experiment for SO 68 is performed as a P.835 listening test.

2.2.2.3.1 Subjective Experiment I for SO 68

The Test Parameters for Listening Experiment I are presented in Table 2.2.2.3.1-1.

Table 2.2.2.3.1-1 SO 68 Listening Experiment I Test Parameters

Condition	Description
Type of test	MOS (P.800)
Number of talkers	4 males, 4 females
Background noise	none (ambient)
Audio Input Level	-22 dB, -32 dB, -12 dB
Filter characteristics	MIRS
Reference conditions	(8) Direct, 3, 9, 15, 21, 27, 33, 39 dBQ

Condition	Description
Test conditions	(a) Low Audio Input Level -32 dB, 9.3, 5.8 kbps, 1% d&b, 1% pls (b) Nominal Audio Input Level, -22 dB, 9.3, 5.8, 4.8 kbps (c) High Audio Input Level -12 dB, 9.3, 5.8 kbps (d) Nominal Audio Input Level, -22 dB, 9.3, 5.8 kbps, 3% FER, M/M, M/T Only
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M: Conditions (a)-(c) (2) M/M, M/T: Condition (d)

The Test Conditions for Listening Experiment I are presented in Table 2.2.2.3.1-2.

Table 2.2.2.3.1-2 SO 68 Listening Experiment I Test Conditions

Label	Operating Point	Condition	Encoder/Decoder Combinations
a01	Reference	MNRU 3dB	
a02	Reference	MNRU 9dB	
a03	Reference	MNRU 15dB	
a04	Reference	MNRU 21dB	
a05	Reference	MNRU 27dB	
a06	Reference	MNRU 33dB	
a07	Reference	MNRU 39dB	
a08	Reference	Direct	
a09	EVRC-B 9.3 kbps	Nominal, -22 dB	M-M
a10	EVRC-B 9.3 kbps	Nominal, -22 dB	M-T
a11	EVRC-B 9.3 kbps	Nominal, -22 dB	T-T
a12	EVRC-B 9.3 kbps	Nominal, -22 dB	T-M
a13	EVRC-B 5.8 kbps	Nominal, -22 dB	M-M
a14	EVRC-B 5.8 kbps	Nominal, -22 dB	M-T
a15	EVRC-B 5.8 kbps	Nominal, -22 dB	T-T
a16	EVRC-B 5.8 kbps	Nominal, -22 dB	T-M
a17	EVRC-B 4.8 kbps	Nominal, -22 dB	M-M
a18	EVRC-B 4.8 kbps	Nominal, -22 dB	M-T
a19	EVRC-B 4.8 kbps	Nominal, -22 dB	T-T
a20	EVRC-B 4.8 kbps	Nominal, -22 dB	T-M
a21	EVRC-B 9.3 kbps	Low, -32 dB, 1% d&b, 1% pls	M-M
a22	EVRC-B 9.3 kbps	Low, -32 dB, 1% d&b, 1% pls	M-T
a23	EVRC-B 9.3 kbps	Low, -32 dB, 1% d&b, 1% pls	T-T
a24	EVRC-B 9.3 kbps	Low, -32 dB, 1% d&b, 1% pls	T-M
a25	EVRC-B 5.8 kbps	Low, -32 dB, 1% d&b, 1% pls	M-M

Label	Operating Point	Condition	Encoder/Decoder Combinations
a26	EVRC-B 5.8 kbps	Low, -32 dB, 1% d&b, 1% pls	M-T
a27	EVRC-B 5.8 kbps	Low, -32 dB, 1% d&b, 1% pls	T-T
a28	EVRC-B 5.8 kbps	Low, -32 dB, 1% d&b, 1% pls	T-M
a29	EVRC-B 9.3 kbps	High, -12 dB	M-M
a30	EVRC-B 9.3 kbps	High, -12 dB	M-T
a31	EVRC-B 9.3 kbps	High, -12 dB	T-T
a32	EVRC-B 9.3 kbps	High, -12 dB	T-M
a33	EVRC-B 5.8 kbps	High, -12 dB	M-M
a34	EVRC-B 5.8 kbps	High, -12 dB	M-T
a35	EVRC-B 5.8 kbps	High, -12 dB	T-T
a36	EVRC-B 5.8 kbps	High, -12 dB	T-M
a37	EVRC-B 9.3 kbps	Nominal, -22 dB, 3% FER	M-M
a38	EVRC-B 9.3 kbps	Nominal, -22 dB, 3% FER	M-T
a39	EVRC-B 5.8 kbps	Nominal, -22 dB, 3% FER	M-M
a40	EVRC-B 5.8 kbps	Nominal, -22 dB, 3% FER	M-T

2.2.2.3.2 Subjective Experiment II for SO 68

The Test Parameters for Listening Experiment II are presented in Table 2.2.2.3.2-1.

Table 2.2.2.3.2-1 SO 68 Listening Experiment II Test Parameters

Condition	Description
Type of test	P-NSA (P.835)
Number of talkers	3 males, 3 females
Background noise	Specified test conditions
Audio Input Level	-22 dB
Filter characteristics	MIRS
Reference conditions	(8) Specified reference conditions
Test conditions	(a) Car Noise @ 15 dB S/N 9.3, 5.8, 4.8 kbps (b) Street Noise @ 15 dB S/N 9.3, 5.8 kbps (c) Office Babble @ 20 dB S/N 9.3, 5.8 kbps
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

The Test Conditions for Listening Experiment II are presented in Table 2.2.2.3.2-2

Table 2.2.2.3.2-2 SO 68 Listening Experiment II Test Conditions

Label	Operating Point	Impairment Condition	Encoder/Decoder Combinations
b01	Reference	Car Noise @ 40 dB SNR, MNRU 40 dB	
b02	Reference	Car Noise @ 20 dB SNR, MNRU 40 dB	
b03	Reference	Car Noise @ 0 dB SNR, MNRU 40 dB	
b04	Reference	Car Noise @ 40 dB SNR, MNRU 0 dB	
b05	Reference	Car Noise @ 40 dB SNR, MNRU 20 dB	
b06	Reference	Car Noise @ 10 dB SNR, MNRU 10 dB	
b07	Reference	Car Noise @ 20 dB SNR, MNRU 20 dB	
b08	Reference	Car Noise @ 30 dB SNR, MNRU 30 dB	
b09	EVRC-B 9.3 kbps	Car Noise @ 15 dB	M-M
b10	EVRC-B 9.3 kbps	Car Noise @ 15 dB	M-T
b11	EVRC-B 9.3 kbps	Car Noise @ 15 dB	T-T
b12	EVRC-B 9.3 kbps	Car Noise @ 15 dB	T-M
b13	EVRC-B 5.8 kbps	Car Noise @ 15 dB	M-M
b14	EVRC-B 5.8 kbps	Car Noise @ 15 dB	M-T
b15	EVRC-B 5.8 kbps	Car Noise @ 15 dB	T-T
b16	EVRC-B 5.8 kbps	Car Noise @ 15 dB	T-M
b17	EVRC-B 4.8 kbps	Car Noise @ 15 dB	M-M
b18	EVRC-B 4.8 kbps	Car Noise @ 15 dB	M-T
b19	EVRC-B 4.8 kbps	Car Noise @ 15 dB	T-T
b20	EVRC-B 4.8 kbps	Car Noise @ 15 dB	T-M
b21	EVRC-B 9.3 kbps	Street Noise @ 15 dB	M-M
b22	EVRC-B 9.3 kbps	Street Noise @ 15 dB	M-T
b23	EVRC-B 9.3 kbps	Street Noise @ 15 dB	T-T
b24	EVRC-B 9.3 kbps	Street Noise @ 15 dB	T-M
b25	EVRC-B 5.8 kbps	Street Noise @ 15 dB	M-M
b26	EVRC-B 5.8 kbps	Street Noise @ 15 dB	M-T
b27	EVRC-B 5.8 kbps	Street Noise @ 15 dB	T-T
b28	EVRC-B 5.8 kbps	Street Noise @ 15 dB	T-M
b29	EVRC-B 9.3 kbps	Office Noise @ 20 dB	M-M
b30	EVRC-B 9.3 kbps	Office Noise @ 20 dB	M-T
b31	EVRC-B 9.3 kbps	Office Noise @ 20 dB	T-T
b32	EVRC-B 9.3 kbps	Office Noise @ 20 dB	T-M
b33	EVRC-B 5.8 kbps	Office Noise @ 20 dB	M-M
b34	EVRC-B 5.8 kbps	Office Noise @ 20 dB	M-T
b35	EVRC-B 5.8 kbps	Office Noise @ 20 dB	T-T
b36	EVRC-B 5.8 kbps	Office Noise @ 20 dB	T-M

2.2.2.3.3 Numerical Parameters for the SO 68 Listening Experiments

Table 2.2.2.3.3-1 describes the resultant numerology that is used for the two SO 68 listening experiments. The first column is a variable name given to each of the parameters, the second

column is the description of the parameter, the third column shows the required calculation for determining the value of the parameter if it is dependent upon other parameter values, and the last two columns show the numerical value for each of the parameters, for the two listening experiments. For each listening experiment, four codecs are evaluated with a differing number of conditions (three for the EVRC-B 9.3 and 6.6 kbps codecs and one for the EVRC-B 5.8 and 4.8 kbps codecs). There are eight reference conditions in both experiments.

Table 2.2.2.3.3-1 Numerical Parameters for the SO 68 Listening Experiments

Var	Parameter	Calculation	Experiment I Value	Experiment II Value
C1	Codecs for Test Condition 1		3	3
C2	Codecs for Test Condition 2		2	2
C3	Codecs for Test Condition 3		2	2
C4	Codecs for Test Condition 4		2	
C5	Codec Combinations: Conditions 1-3		4	4
C6	Codec Combinations: Condition 4		2	
C7	Reference Conditions		8	8
C8	Total Conditions	(C1*C2+C3)*C5+ C4*C6+C7	40	36
C9	Talkers		8	6
C10	Stimuli per Talker		8	8
C11	Stimuli per Condition	C9*C10	64	48
C12	Total Stimuli per Experiment	C8*C11	2560	1728
C13	Listening Panels		8	8
C14	Stimuli per Listening Panel	C8*C9	320	216
C15	Listeners (Voters)		32	32
C16	Listeners (Voters) per Listening Panel	C15/C13	4	4
C17	Votes per Condition	C9*C13	256	192

2.2.3 Speech Material for SO 68 Testing

The source speech files used for SO 68 compliance testing consist of 128 Harvard sentences, which are preprocessed to include proper level adjustment and noise mixing for use in the two subjective experiments. The talkers used in these files consist of four adult males and four adult females, and are native speakers of North American English.

For the following discussion, it may be useful to refer to Table 4-2 for the composition of the Software Distribution database.

2.2.3.1 Source Speech Material for SO 68 Experiment I

The source speech material for subjective Experiment I is contained in directory /so68/subjctv/exp1/source. Each file is MIRS filtered and level adjusted to -22, -12, or -32 dB. These files are named src.s22, src.s12 and src.s32, respectively. The speech database also includes samples processed through the various reference conditions in directory /so68/subjctv/exp1/ref. The reference conditions are named a01 through a08 for the respective conditions given in Table 2.2.2.3.1-2.

2.2.3.2 Source Speech Material for SO 68 Experiment II

The source speech material for subjective Experiment II is contained in directory /so68/subjctv/exp2/source. This directory contains the source material for the car, street, and babble noise conditions, which are named src.c15, src.s15, and src.b20, respectively. The speech database also includes samples processed through the various reference conditions in directory /so68/subjctv/exp2/ref. The reference conditions are named ref.b01 through ref.b08 for the respective conditions given in Table 2.2.2.3.2-2.

2.2.4 Processing of Speech Material for SO 68 Testing

The source speech material shall be processed by the various combinations of encoders and decoders listed in the descriptions of the two experiments given in Section 2.2.2. The master codec software described in Section 3.2.3 shall be used in the processing involving the master codec. Generally, the master codec encoder and decoder outputs have been provided in the respective directories, /so68/subjctv/exp*/m_pkt and /so68/subjctv/exp*/m_m. Execution of the master codec software is needed only for the test encoder/master decoder combination for each experiment/condition.

All codec processing shall be done digitally. Noise suppression and post-filter options shall be enabled for both the master and the test codecs. The digital format of the speech files is described in Section 3.2.4.4.

The naming convention of the processed speech is as follows: For the packet files in the /so68/subjctv/exp1/m_pkt directory (Experiment I), the *.p12 files are the master packet files for the *.s12 source file. Likewise, the *.p22 and *.p32 files are the respective packet files for the *.s22 and *.s32 source files. For the packet files, the file name 9_3.* indicates an output from the master encoder at 9.3 kbps active speech channel rate. Likewise, the file names 5_8,* and 4_8.* indicate an output from the master encoder at the respective active speech channel rates. The *.pf3 files are the impaired packet files which will be described in Section 2.2.4.3.

Similarly, the directory /so68/subjctv/exp2/m_pkt contains the master packet files for Experiment II. Here, the *.pc, *.pb, and *.ps files are the master packet files for the *.c15, *.b20, and *.s15 source files, respectively.

For the master encode/master decode directories (/so68/subjctv/exp*/m_m), the naming convention of the speech files is such that the first two characters of the file name indicate the codec combination and the suffix indicates the condition numbers in Table 2.2.2.3.1-2 and Table 2.2.2.3.2-2.

Detailed descriptions of all processing operations are given in Section 6 .

2.2.4.1 Encoding by the Test Codec

All of the source files will be encoded by the test codec to produce encoded packet files. For ease of reference, it is recommended that directories /so68/subjctv/exp1/t_pkt and /so68/subjctv/exp2/t_pkt be created to deposit the test encoder output packets, and that the naming conventions be made consistent with the master codec.

2.2.4.2 Decoding by the Master/Test Codecs

The encoded packet files generated from the various encoders/conditions shall be processed through the master and test decoders.

2.2.4.3 Introduction of Impairments

For the 3% frame error condition (Experiment I, condition (d)), the impaired master codec encoded packet files are provided in the /so68/subjctv/exp1/m_pkt directory. Unlike other conditions, this condition uses only the test decoder and not the test encoder.

For the Dim-and-Burst processing, and also the Packet Level Signaling conditions in Experiment I, the processing requires inputs from a signaling file to control maximum encoding rate. An external software utility (EvrcB_iwf in Section 3.2.2.3) is also needed to reduce the data rate of certain packets from full rate to half rate. Details of these operations are given in Section 6. The signaling file and other utilities are provided in /so68/tools/ directory.

2.2.4.4 Ensuring Proper Encoded Frame Packet Files

All encoded frame packet files shall be examined to ensure that the files only contain data in those file locations where data should exist for a given data rate.

The examination of the encoded frame packet files should indicate the occurrence of any improper data in the files but the examination must not alter the encoded frame packet files in any way.

2.2.4.5 Post-processing of test-condition output files

In order to build the play sets to be presented to the listening panels the output files for the various test conditions must be processed to provide the appropriate listening conditions. In addition, the concatenated output files must be partitioned into the samples representing the combination of test-condition and talker. The listening conditions are provided by filtering the output files using the STL software tool (*filter*) with the MIRS-receive filter mask. An STL tool (*astrip*) is also used to split the concatenated files into the individual samples appropriate for the experiment. Table 2.2.4.5-1 shows the cutting-points to be used with the *astrip* tool for producing the two-sentence samples for the Experiment I ACR test. Table 2.2.4.5-2 shows the cutting-points to be used with the *astrip* tool for producing the single-sentence sub-samples for the Experiment II P.835 test. Table 2.2.4.5-3 shows the sub-samples that make up the samples (i.e., sentence triads) for the P.835 test.

Table 2.2.4.5-1 Cutting Points for the astrip Software Tool for the Experiment I ACR Test

Experiment I - ACR							
Sentence-pair		Start Sample	Length (samples)	Sentence-pair		Start Sample	Length (samples)
1	m1p1	1	49664	33	m1p5	1636353	49664
2	f1p1	49665	49152	34	f1p5	1686017	50432
3	m2p1	98817	53504	35	m2p5	1736449	50176
4	f2p1	152321	57600	36	f2p5	1786625	55296
5	m3p1	209921	47616	37	m3p5	1841921	50944
6	f3p1	257537	47360	38	f3p5	1892865	48384
7	m4p1	304897	52736	39	m4p5	1941249	54784
8	f4p1	357633	51712	40	f4p5	1996033	54016
9	m1p2	409345	50688	41	m1p6	2050049	50432
10	f1p2	460033	50176	42	f1p6	2100481	50688
11	m2p2	510209	53504	43	m2p6	2151169	56320
12	f2p2	563713	50944	44	f2p6	2207489	51712
13	m3p2	614657	51456	45	m3p6	2259201	56576
14	f3p2	666113	48128	46	f3p6	2315777	45824
15	m4p2	714241	51712	47	m4p6	2361601	52480
16	f4p2	765953	49920	48	f4p6	2414081	50944
17	m1p3	815873	49408	49	m1p7	2465025	53760
18	f1p3	865281	45568	50	f1p7	2518785	49152
19	m2p3	910849	50176	51	m2p7	2567937	47360
20	f2p3	961025	51968	52	f2p7	2615297	57088
21	m3p3	1012993	54016	53	m3p7	2672385	54784
22	f3p3	1067009	49408	54	f3p7	2727169	45568
23	m4p3	1116417	53760	55	m4p7	2772737	51200
24	f4p3	1170177	51968	56	f4p7	2823937	52736
25	m1p4	1222145	47104	57	m1p8	2876673	49408
26	f1p4	1269249	47104	58	f1p8	2926081	47616
27	m2p4	1316353	50944	59	m2p8	2973697	55808
28	f2p4	1367297	54272	60	f2p8	3029505	54272
29	m3p4	1421569	53248	61	m3p8	3083777	63232
30	f3p4	1474817	50432	62	f3p8	3147009	46336
31	m4p4	1525249	56320	63	m4p8	3193345	55040
32	f4p4	1581569	54784	64	f4p8	3248385	50176

Table 2.2.4.5-2 Cutting Points for the *astrip* Software Tool for the Experiment II P.835 Test

Experiment II - P.835												
	Sentence	Start Sample	Length (samples)		Sentence	Start Sample	Length (samples)		Sentence	Start Sample	Length (samples)
1	m1s01	1	25242	33	m3s05	1012993	23995	65	m2s11	2151169	23796
2	m1s02	25243	24422	34	m3s06	1036988	30021	66	m2s12	2174965	32524
3	f1s01	49665	21072	35	f3s05	1067009	20723	67	f2s11	2207489	23719
4	f1s02	70737	28080	36	f3s06	1087732	28685	68	f2s12	2231208	27993
5	m2s01	98817	27194	37	m1s07	1222145	21654	69	m3s11	2259201	23729
6	m2s02	126011	26310	38	m1s08	1243799	25450	70	m3s12	2282930	32847
7	f2s01	152321	26955	39	f1s07	1269249	23163	71	f3s11	2315777	20687
8	f2s02	179276	30645	40	f1s08	1292412	23941	72	f3s12	2336464	25137
9	m3s01	209921	21939	41	m2s07	1316353	21946	73	m1s13	2465025	23992
10	m3s02	231860	25677	42	m2s08	1338299	28998	74	m1s14	2489017	29768
11	f3s01	257537	22946	43	f2s07	1367297	27136	75	f1s13	2518785	23256
12	f3s02	280483	24414	44	f2s08	1394433	27136	76	f1s14	2542041	25896
13	m1s03	409345	23249	45	m3s07	1421569	26239	77	m2s13	2567937	23386
14	m1s04	432594	27439	46	m3s08	1447808	27009	78	m2s14	2591323	23974
15	f1s03	460033	20319	47	f3s07	1474817	24122	79	f2s13	2615297	28367
16	f1s04	480352	29857	48	f3s08	1498939	26310	80	f2s14	2643664	28721
17	m2s03	510209	24265	49	m1s09	1636353	21087	81	m3s13	2672385	26883
18	m2s04	534474	29239	50	m1s10	1657440	28577	82	m3s14	2699268	27901
19	f2s03	563713	25104	51	f1s09	1686017	25112	83	f3s13	2727169	19206
20	f2s04	588817	25840	52	f1s10	1711129	25320	84	f3s14	2746375	26362
21	m3s03	614657	22326	53	m2s09	1736449	22289	85	m1s15	2876673	23122
22	m3s04	636983	29130	54	m2s10	1758738	27887	86	m1s16	2899795	26286
23	f3s03	666113	20484	55	f2s09	1786625	26163	87	f1s15	2926081	20020
24	f3s04	686597	27644	56	f2s10	1812788	29133	88	f1s16	2946101	27596
25	m1s05	815873	22969	57	m3s09	1841921	25367	89	m2s15	2973697	25310
26	m1s06	838842	26439	58	m3s10	1867288	25577	90	m2s16	2999007	30498
27	f1s05	865281	23114	59	f3s09	1892865	21843	91	f2s15	3029505	26239
28	f1s06	888395	22454	60	f3s10	1914708	26541	92	f2s16	3055744	28033
29	m2s05	910849	24362	61	m1s11	2050049	22924	93	m3s15	3083777	27501
30	m2s06	935211	25814	62	m1s12	2072973	27508	94	m3s16	3111278	35731
31	f2s05	961025	25286	63	f1s11	2100481	23930	95	f3s15	3147009	20918
32	f2s06	986311	26682	64	f1s12	2124411	26758	96	f3s16	3167927	25418

Table 2.2.4.5-3 Composition of the Sentence-Triad Samples for the Experiment II P.835 Test

Sentence-triad	Sentence 1	Sentence 2	Sentence 3
t1	s01	s02	s03
t2	s04	s05	s06
t3	s07	s08	s09
t4	s10	s11	s12
t5	s13	s14	s15
t6	s16	s01	s02
t7	s03	s04	s05
t8	s06	s07	s08

2.2.5 Randomization

For each of the two subjective experiments, each presentation sample consists of a speech sample processed under a condition of the test. For the ACR Experiment I the sample consists of a pair of concatenated sentences of approximately 8 sec. duration. For the P.835 Experiment II the sample consists of three sub-samples, where each sub-sample is a single sentence of approximately 4 sec. duration. The samples shall be presented to the listeners in a randomized presentation order. The listeners for each file set shall be presented with practice trials for subjective Experiments I and II. The randomization of the test samples has been accomplished with the following constraints for the two experiments:

1. A trial, i.e., a test sample, for the combination of each test condition and each talker shall be presented exactly once to each listening panel (i.e., # trials/panel = # conditions x # talkers).

2. Randomization is in “blocks”, such that one sample of each test condition is presented once, with a randomly selected talker, in each block. This ensures that listeners rate each test condition equally often in the initial, middle and final parts of the block and controls for the effects of time and order of presentation. A block contains the same number of samples as there are test-conditions involved in the test. A test “session” consists of the same number of blocks as there are talkers involved in the test. Each session is presented to a listening panel of four listeners.

3. Randomizations are constructed such that talker gender is alternated on successive trials resulting in the same talker never being presented on consecutive trials.

Table 2.2.5-1 shows an example randomization for a single listening panel. Each entry in the table is the file name for a sample with the following file-naming convention - ***xxyy.zzz***, where ***xx*** is the talker, ***yy*** is the sample, and ***zzz*** is the test condition.

Table 2.2.5-1 Example Randomization for the Experiment I ACR Test

Panel 1	Blk 1	Blk 2	Blk 3	Blk 4	Blk 5	Blk 6	Blk 7	Blk 8
1	f2p8.a06	f4p8.a14	f1p3.a25	f4p5.a17	f2p5.a01	f1p8.a12	f4p7.a34	f3p5.a20
2	m3p8.a03	m2p1.a24	m4p1.a24	m1p6.a18	m2p3.a29	m4p4.a27	m3p1.a13	m3p7.a16
3	f2p7.a22	f4p8.a06	f2p4.a34	f3p1.a16	f2p6.a33	f3p3.a38	f1p7.a07	f3p3.a12
4	m1p6.a09	m1p3.a23	m1p8.a21	m2p3.a11	m1p2.a20	m3p4.a02	m2p8.a20	m4p4.a09
5	f3p2.a07	f2p3.a36	f3p2.a27	f1p1.a14	f3p4.a34	f1p2.a20	f1p4.a39	f4p8.a05
6	m3p8.a19	m3p2.a17	m1p7.a29	m3p8.a04	m4p1.a31	m3p7.a18	m2p2.a36	m1p4.a14
7	f4p6.a16	f2p4.a04	f2p7.a10	f2p6.a07	f1p6.a08	f3p2.a14	f3p7.a09	f2p4.a03
8	m2p8.a34	m2p6.a32	m3p5.a31	m3p7.a12	m3p3.a06	m1p8.a40	m2p8.a04	m4p8.a33
9	f3p8.a39	f2p2.a28	f1p2.a09	f4p4.a33	f1p1.a16	f2p7.a37	f4p2.a26	f4p3.a37
10	m4p1.a28	m1p8.a07	m4p1.a16	m2p8.a03	m3p3.a38	m4p8.a19	m3p5.a29	m4p1.a25
11	f1p3.a05	f1p7.a35	f2p6.a18	f2p8.a31	f1p7.a32	f3p2.a22	f2p8.a16	f2p3.a35
12	m4p1.a12	m2p8.a40	m2p3.a38	m2p7.a35	m4p3.a07	m2p1.a33	m3p3.a05	m4p5.a17
13	f1p3.a37	f4p6.a22	f1p4.a33	f3p8.a08	f1p6.a40	f2p2.a29	f2p4.a08	f2p7.a19
14	m4p6.a20	m3p7.a33	m3p8.a23	m2p3.a19	m2p5.a21	m3p2.a26	m1p6.a19	m1p3.a30
15	f3p2.a23	f3p7.a21	f4p2.a04	f4p2.a09	f4p7.a27	f1p2.a36	f4p4.a02	f1p8.a18
16	m3p4.a27	m4p2.a34	m2p6.a14	m2p1.a27	m1p4.a28	m2p1.a09	m4p4.a22	m3p8.a40
17	f2p6.a30	f1p2.a27	f1p5.a01	f1p6.a30	f4p3.a03	f4p5.a31	f3p7.a17	f4p3.a13
18	m2p2.a26	m4p6.a18	m1p5.a13	m4p1.a29	m3p1.a22	m4p5.a03	m4p5.a06	m3p6.a24
19	f1p7.a29	f1p3.a03	f4p3.a36	f3p1.a32	f2p7.a09	f3p6.a06	f2p7.a24	f1p2.a02
20	m2p8.a18	m3p5.a01	m4p8.a32	m4p6.a37	m4p8.a15	m2p5.a17	m3p7.a21	m2p5.a23
21	f1p1.a21	f1p3.a11	f4p3.a12	f2p3.a39	f2p2.a25	f1p4.a04	f1p7.a15	f4p7.a29
22	m4p1.a04	m2p8.a16	m3p5.a15	m4p5.a05	m2p1.a37	m2p8.a01	m4p8.a38	m2p2.a39
23	f2p2.a38	f4p1.a38	f2p7.a26	f1p5.a38	f3p2.a18	f2p6.a21	f3p8.a33	f3p3.a36
24	m1p3.a17	m4p7.a10	m2p2.a30	m1p7.a02	m4p8.a39	m3p3.a34	m4p2.a14	m3p3.a32
25	f2p3.a14	f2p5.a12	f3p1.a19	f1p6.a06	f4p4.a35	f2p1.a05	f3p4.a01	f2p3.a11
26	m1p4.a01	m2p8.a08	m1p6.a05	m4p8.a21	m1p1.a12	m3p8.a10	m3p6.a37	m4p3.a01
27	f3p7.a15	f1p4.a19	f3p8.a03	f3p3.a40	f3p4.a10	f4p2.a23	f1p3.a23	f3p5.a28
28	m3p3.a35	m4p4.a02	m2p6.a22	m4p2.a13	m3p2.a30	m1p5.a08	m1p5.a03	m1p6.a38
29	f4p7.a08	f3p4.a37	f3p2.a11	f4p3.a25	f3p8.a02	f1p5.a28	f2p1.a40	f1p2.a26
30	m3p5.a11	m1p4.a39	m4p1.a40	m3p4.a36	m2p6.a13	m1p3.a16	m2p3.a28	m2p5.a07
31	f3p8.a31	f4p3.a30	f4p1.a20	f4p6.a01	f4p7.a19	f3p7.a30	f3p7.a25	f3p3.a04
32	m1p2.a33	m1p8.a15	m4p7.a08	m1p3.a10	m2p4.a05	m4p6.a11	m1p8.a11	m3p3.a08
33	f4p7.a40	f3p7.a05	f3p4.a35	f3p6.a24	f2p6.a17	f4p6.a07	f1p4.a31	f1p7.a34
34	m1p4.a25	m3p8.a25	m1p5.a37	m1p1.a26	m4p7.a23	m4p4.a35	m1p8.a35	m2p7.a31
35	f4p8.a32	f3p4.a13	f4p6.a28	f2p6.a23	f4p1.a11	f4p4.a15	f2p1.a32	f2p1.a27
36	m2p8.a02	m1p4.a31	m2p6.a06	m3p1.a20	m1p7.a04	m1p6.a32	m2p8.a12	m2p2.a15
37	f4p3.a24	f2p6.a20	f1p1.a17	f1p3.a22	f3p4.a26	f4p4.a39	f4p5.a18	f4p8.a21
38	m2p8.a10	m4p4.a26	m3p1.a07	m1p7.a34	m1p8.a36	m2p7.a25	m4p6.a30	m1p3.a22
39	f1p1.a13	f3p4.a29	f2p8.a02	f2p3.a15	f1p3.a24	f2p1.a13	f4p7.a10	f1p6.a10
40	m4p6.a36	m3p4.a09	m3p5.a39	m3p7.a28	m3p4.a14	m1p8.a24	m1p8.a27	m1p4.a06

The randomization lists for each of the eight listening panels for each experiment are provided in /so68/subjctv/exp1/data/play*.lst and /so68/subjctv/exp2/data/play*.lst, respectively.

2.2.6 Presentation

Presentation of speech materials for the SO 68 codec listening tests shall be made with one side of high fidelity supra-aural headphones with the other ear uncovered. The speech material delivery system shall meet the requirements of Section 3.2.1.1. The listeners should be seated in a quiet room, with an ambient noise level of 30 dBA or below.

2.2.7 Listeners

The listener sample is intended to represent the population of telephone users with normal hearing acuity. The listeners should be naïve with respect to telephony technology issues; that is, they should not be experts in telephone design, digital voice encoding algorithms, and so on. They should not be trained listeners; that is, they should not have been trained in these or previous listening studies using feedback trials. Age distribution and gender should be nominally balanced across listening panels.

Each listener shall provide data only once for a particular evaluation. A listener may participate in different evaluations, but test sessions performed with the same listener should be at least two months apart so as to reduce the cumulative effects of experience.

2.2.8 Listening Test Procedures

2.2.8.1 ACR Listening Test Procedures – Experiment I.

The listeners shall listen to each sample and rate the quality of the test sample using a five-point scale, with the points labeled:

5 Excellent

4 Good

3 Fair

2 Poor

1 Bad

Data from 32 listeners shall be used for Experiment I, four listeners for each listening panel where each listening panel uses a different randomization. Before starting the test, the listeners should be given instructions for performing the subjective test. An example set of instructions for the ACR are presented in Figure 2.1.8-1. The instructions may be modified to allow for variations in laboratory data-gathering apparatus.

This is an experiment to determine the perceived quality of speech over the telephone. You will be listening to a number of recorded speech samples, spoken by several different talkers, and you will be rating how good you think they sound.

Use the single headphone on the ear you normally use for the telephone. On each trial a two-sentence sample will be played. After you have listened to the sample, determine the category from the list below which best describes the overall quality of the sample. Press the numeric key on your keyboard corresponding to your rating for how good or bad that particular passage sounded.

The quality of the speech should be rated according to the scale below:

5 Excellent

4 Good

3 Fair

2 Poor

1 Bad

During the session you will hear samples varying in different aspects of quality. Please take into account your total impression of each sample, rather than concentrating on any particular aspect.

Figure 2.2.8.1-1 Instructions for Listeners

2.2.8.2 P-835 Listening Test Procedures – Experiment II

Experimental II uses the P.835 test methodology described in ITU-T Rec. P.835 [13]. The P.835 methodology is specifically designed to evaluate the quality of speech in background noise. It yields a measure of Signal Quality (SIG), a measure of Background Quality (BAK), and a measure of Overall Quality (OVRL). In general, OVRL scores are highly correlated with MOS but the OVRL score provides greater sensitivity and precision in test conditions involving background noise. While the OVRL score is of most interest here, the SIG and BAK scores also provide valuable diagnostic information. For each trial in a P.835 test, listeners are presented with three sub-samples where each sub-sample is a single sentence (approx. 4 sec. duration) processed through the same test condition. In one of the first two sub-samples listeners rate the *Signal Quality* on a five-point rating scale with the points labeled:

5 Very natural, no distortion
4 Fairly natural, little distortion
3 Somewhat natural, some distortion
2 Fairly unnatural, fairly distorted
1 Very unnatural, very distorted

For the other of the first two sub-samples listeners rate the *Background Quality* on a five-point rating scale with the points labeled:

5	Not noticeable
4	Fairly noticeable
3	Noticeable but not intrusive
2	Fairly conspicuous, somewhat intrusive
1	Very conspicuous, very intrusive

For the third sub-sample listeners rate the *Overall quality* on a five-point rating scale with the points labeled:

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Data from 32 listeners shall be used for Experiment II, four listeners for each listening panel where each listening panel uses a different randomization

Before starting the test, the listeners should be given instructions for performing the subjective test. An example set of instructions for the P.835 test are presented below. The instructions may be modified to allow for variations in laboratory data-gathering apparatus.

Instructions for P.835 Speech Rating Experiment

In this speech rating experiment each trial will involve **three** sentences and you will give a rating for **each** sentence.

For the first sentence in each trial you will be asked to attend **only to the speech signal** and rate how natural, or conversely, how degraded, the **speech signal** sounds to you. You will use the rating scale shown in the figure below to register your ratings of the speech signal. Your task will be to choose the numbered phrase from the list below that best describes your opinion of the **SPEECH SIGNAL ALONE** and then enter the corresponding number on your keyboard.

Attending **ONLY to the SPEECH SIGNAL**, select the category which best describes the sample you just heard.

the **SPEECH SIGNAL** in this sample was

5 – VERY NATURAL, NO DEGRADATION
4 – FAIRLY NATURAL, LITTLE DEGRADATION
3 - SOMEWHAT NATURAL, SOMEWHAT DEGRADED
2 – FAIRLY UNNATURAL, FAIRLY DEGRADED
1 – VERY UNNATURAL, VERY DEGRADED

For the second sentence in each trial you will be asked to attend **only to the background** and rate how noticeable, intrusive, and/or conspicuous the **background** sounds to you. You will use the rating scale shown in the figure below to register your ratings of the background. Your task will be to choose the numbered phrase from the list below that best describes your opinion of the **BACKGROUND ALONE** and then enter the corresponding number on your keyboard.

Attending **ONLY to the BACKGROUND**, select the category which best describes the sample you just heard.

the **BACKGROUND** in this sample was

5 – NOT NOTICEABLE
4 – SOMEWHAT NOTICEABLE
3 – NOTICEABLE BUT NOT INTRUSIVE
2 – FAIRLY CONSPICUOUS, SOMEWHAT INTRUSIVE
1 – VERY CONSPICUOUS, VERY INTRUSIVE

For the third and final sentence in each trial you will be asked to attend to the entire sample (both the speech signal and the background) and rate your opinion of the sample for purposes of everyday speech communication.

Select the category which best describes the sample you just heard for purposes of everyday speech communication.

the **OVERALL SPEECH SAMPLE** was

5 – EXCELLENT
4 – GOOD
3 - FAIR
2 – POOR
1 – BAD

2.2.9 Analysis of Results

The response data from the practice blocks shall be discarded. Data sets with missing responses from listeners shall not be used – i.e., a complete set of data is required for 32 listeners, four for each of eight listening panels. Responses from the different listening panels for the corresponding test conditions shall be treated as equivalent in the analysis.

2.2.9.1 Basic Results for the SO 68 Listening tests

The votes for each of the test conditions for SO 68 Experiments I and II shall be averaged to produce an associated mean score (***M***) as shown in Equation 2.2.9.1-1 and a Standard Deviation (***SD***) as shown in Equation 2.2.9.1-2, where ***L*** is the number of listeners and ***T*** is the number of talkers involved in the experiment.

$$\overline{M} = \left(\sum_{L} \sum_{T} X_{l,t} \right) \Big/ \left(L \times T \right) \qquad (2.2.9.1\text{-}1)$$

$$SD = \sqrt{\left(\sum_{L} \sum_{T} \left(X_{l,t} - \overline{M} \right)^2 \right) \Big/ \left(L \times T - 1 \right)} \qquad (2.2.9.1\text{-}2)$$

2.2.9.2 Minimum Subjective Requirement for SO 68 Listening Tests

The Terms of Reference for the MPS tests state that the mean score for each of the Test Encoder/Decoder Combinations (E/DC) should be "*not worse than*" the mean score for the Reference E/DC. For most of the test conditions involved in the subjective experiments there are three Test E/DC's (M-T, T-M, and T-T) which means there are three statistical tests against the Reference E/DC (M-M). The three statistical tests are not independent, however. Since they all involve the same ratings for the Reference E/DC, t-tests are not appropriate. The appropriate statistical test for multiple Test conditions against a common Reference condition is Dunnett's Test. A complete description of Dunnett's Test is contained in Appendix B.

The critical value for the Dunnett's test is 2.09 (one-sided test, $p<.05$, 4 E/DC's, df = 93).

For those test conditions where a single Test E/DC (T-T) is compared against the Reference E/DC (M-M), the appropriate statistical test is Student's t-test[4].

The critical value for the Student's t-test is 1.70 (one-sided test, $p<.05$, df = 31).

In both the Dunnett's Test and the t-test the MPS test is evaluated by dividing the difference between the mean score for the Test E/DC and the mean score for the Reference ED/C by the Standard Error of the Mean Difference (SE_{MD}) as shown in Equation 2.2.9.2-1. If the resultant Test value is less than the criterion value for the appropriate test (2.09 for Dunnett's Test, 1.70 for the t-test), then the E/DC passes the MPS test.

---

[4] The appropriate t-test is a "matched groups" t-test and the SE_{MD} is based on the differences between individual listener's average ratings, where the average is over talkers. Therefore, the SE_{MD} is based on 32 difference scores, one for each listener (df =31).

$$Test = \left(\overline{M}_{Ref} - \overline{M}_{Test}\right) \Big/ SE_{MD} \qquad (2.2.9.2\text{-}1)$$

2.2.10 Expected Results for Reference Conditions

2.2.10.1 Experiment I Reference Conditions

The MNRU conditions have been included to provide a frame of reference for the Experiment I MOS test. In listening evaluations where test conditions span approximately the same range of quality, the MOS results for similar conditions should be approximately the same. Data from previous studies allows a generalization to be made concerning the expected MOS results for the MNRU reference conditions (see Figure 2.2.10.1-1).

MOS scores obtained for the MNRU conditions in any SO 68 validation test should be compared to those shown in the graph below. Inconsistencies beyond a small shift in the means in either direction or a slight stretching or compression of the scale near the extremes may imply a problem in the execution of the evaluation test. In particular, MOS should be monotonic with MNRU, within the limits of statistical resolution; and the contour of the relation should show a similar slope.

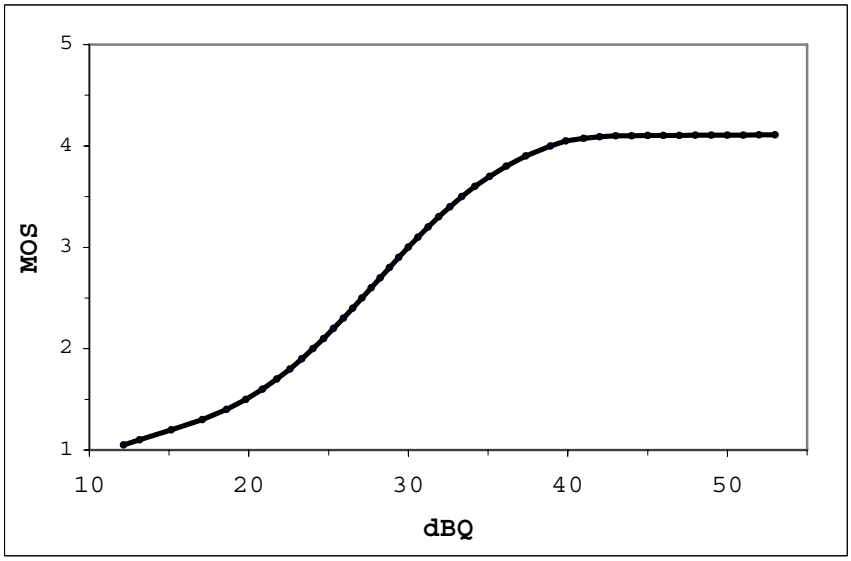


Figure 2.2.10.1-1 MOS versus MNRU

2.2.10.2 Experiment II Reference Conditions

Reference conditions for P.835 tests are constructed as a combination of SNR and MNRU processing to provide degradation in overall speech quality in two dimensions — signal distortion and background noise intrusiveness. Table 2.2.2.3.2-2 shows the eight reference conditions (b01 – b08) involved in the P.835 Experiment II. In general, results are expected for these reference conditions such that the obtained score profiles are similar to those shown in Figure 2.2.10.2-1.

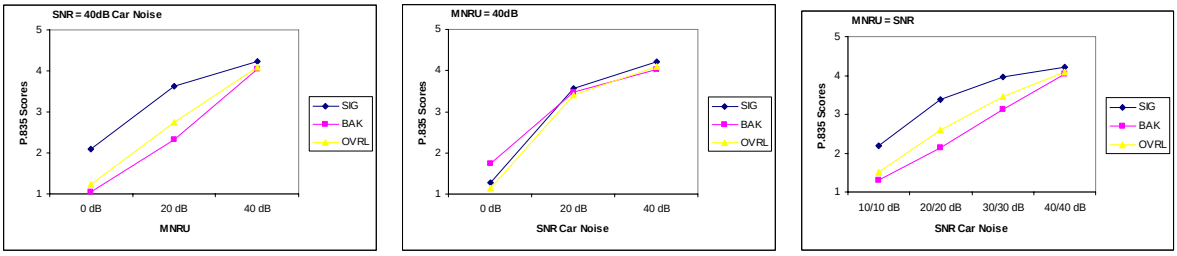


Figure 2.2.10.2-1 P.835 Score Profiles for Reference Conditions

2.3 Performance Testing for SO 70

2.3.1 Objective Performance Testing for SO 70

The objective testing portion of this specification consists of an average data rate test, and compliance to End-to-End Algorithmic Delay and Unity-gain requirements.

2.3.1.1 Average Data Rate Test

An implementation may support SO 70 for 16 kHz sample rates (for example, a mobile station that supports wideband electro-acoustics). The average data rate for the test codec shall be measured using seven source speech files that are contained in the /so70/subjctv/exp{1,2}/source/ directories. Each file exhibits a different condition: power levels: -12 dB, -22 dB, and -32 dB, and background noise conditions: 20 dB SNR babble noise, 10 dB SNR car noise, 20 dB SNR car noise and 15 dB SNR street noise. The input source files used in the average data rate test have an approximate voice activity factor of 0.6, and are the same input files used in the subjective portion of the experiment.

An implementation may support SO 70 only for 8 kHz sample rate input/output (for example, a Base-station transcoder or a Media Gateway or mobile station). The average data rate for the test codec for this case shall be measured using six narrow-band source speech files that are contained in the /so70/subjctv/exp{3,4}/source/ directories. Each file exhibits a different condition: power levels: -12 dB, -22 dB, and -32 dB, and background noise conditions: 20 dB SNR babble noise, 15 dB SNR car noise and 15 dB SNR street noise. The input source files used in the average data rate test have an approximate voice activity factor of 0.6, and are the same input files used in the subjective portion of the experiment.

2.3.1.1.1 Average Data Rate Computation for SO 70

The average channel data rate for the test codec shall be computed for each of the benchmark files as follows:

$$R = (9600*N_1 + 4800*N_2 + 1200*N_8)/N,$$

where

N_1 = number of frames encoded at Rate 1,

N_2 = number of frames encoded at Rate 1/2,

N_8 = number of frames encoded at Rate 1/8, and

$N = N_1 + N_2 + N_8$.

For the 16 kHz input :

The total average channel data rate for the test codec is then given by:

R_{avg} = 1/7 * { R(ambient background segment @ -12dB).+ R(ambient background segment @ -32dB).+ R(ambient background segment @ -22dB). + R(20 dB SNR babble noise segment @ -22dB) + R(10 dB SNR car noise segment @ -22dB). + R(20 dB SNR car noise segment @ -22dB).+ R(15 dB SNR street noise segment @ -22dB) }.

For the 8 kHz input :

The total average channel data rate for the test codec is then given by:

R_{avg} = 1/6 * { R(ambient background segment @ -12dB).+ R(ambient background segment @ -32dB).+ R(ambient background segment @ -22dB). + R(20 dB SNR babble noise segment @ -22dB) + R(15 dB SNR car noise segment @ -22dB). + R(15 dB SNR street noise segment @ -22dB)

The above files are to be processed with EVRC-WB encoder at various capacity operating points (defined by the active speech average channel rate) shown in Table 2.3.1.1.1-1.

Table 2.3.1.1.1-1 Target ADR vs Capacity Operating Point

Capacity Operating Point (active speech average channel data rate)	Target Average Channel Data Rate, kbps
EVRC-WB - RATE_REDUC='000'	5.6+1.5%
EVRC-WB - RATE_REDUC='100'	5.92+1.5%
EVRC-WB - RATE_REDUC='111'	3.29+1.5%

The above table provides the maximum allowable average channel rate (including full, half, and eighth-rate) for the different operating points. These maximum allowable average channel rates were obtained by processing the 7 wide-band benchmark files for the 16 kHz case and 6 narrow-band benchmark files for the 8kHz case through the master floating point software. See Section 3.3.2.1 for details in using the provided software tool that can be used to aid in making this calculation.

2.3.1.1.2 Average Data Rate Requirement for SO 70

The total average data rate R_{avg} for each operating point shall not exceed the target average data rate by more than the tolerance level in Table 2.3.1.1.1-1, otherwise the test codec fails the compliance test.

2.3.1.2 Unity Gain Requirement

The specific EVRC-WB test codec shall output speech with unity gain when compared with the input speech. The unity gain measurement (output active–speech level/input active speech level) will be performed over the entire input speech database for the clean, nominal-level source conditions for each mode. The measurement should be made using the STL-2000 tool [6] [6a] actlev, and must not show more than ± 0.5 dB deviation between input and output active speech levels. This procedure is fully described in [9].

2.3.1.3 End-to-end Algorithmic Delay Recommendation

The algorithmic delay for the specific EVRC-WB test codec should be calculated analytically by the codec manufacturer. In considering the algorithmic delay, it can be assumed that all transmission channels have infinite bandwidth, and that all processing elements have infinite throughput. Algorithmic delay is defined as the sum of all sequential filter delays and buffering delays in the encode/decode path.

The maximum end-to-end algorithmic delay should be no greater than that of the master codec. For the master codecs defined in [1], the algorithmic delay is given as:

Delay Source	Delay (ms)
Signal Preprocessing Delay:	0.0
Filterbank Analysis	0.8
LPC Analysis “Look-ahead”:	10.0
LPC Analysis Window:	20.0
Highband excitation generation delay	1.5
Highband synthesis overlap-and-add delay	2.0
Filterbank Synthesis Delay:	1.1
Total:	35.4

Therefore, the total algorithmic delay imposed by a SO 70 test codec should not exceed 35.4 milliseconds.

2.3.2 Subjective Performance Testing for SO 70

This section outlines the subjective testing methodology of the subjective performance test. The purpose of this testing is to evaluate the quality of the test codec under a variety of conditions which may occur in the CDMA system. To accomplish this, suites of listening experiments have been designed to test speech codec quality under a variety of conditions depending on a number of parameters. These conditions include channel impairments, audio background noise, and different input levels.

Figure 2.3.2-1 illustrates a decision tree to arrive at the suite of tests that are needed to demonstrate Minimum Performance Spec compliance of a Test implementation of SO 70 for different profiles of equipment that support SO 70.

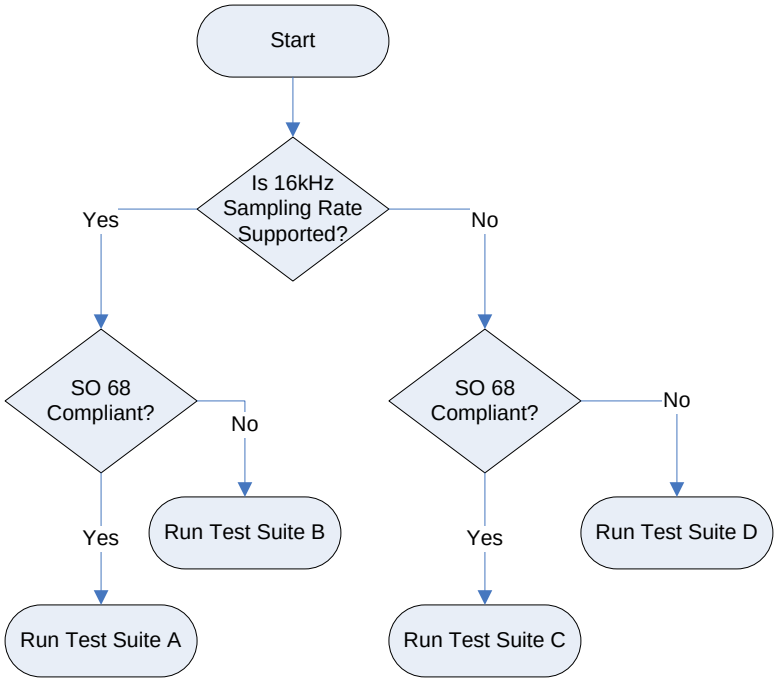


Figure 2.3.2-1 SO 70 Subjective test suite decision flowchart

An implementation may support SO 70 only for 8 kHz sample rate input/output (for example, a Base-station transcoder or a Media Gateway). An implementation may support SO 70 for both 16 kHz and 8 kHz sample rates (for example, a mobile station that supports wideband electro-acoustics).

Further, the implementation supporting SO 70 might already have demonstrated compliance to SO 68 Minimum Performance Spec. This means that such equipment has also demonstrated the Minimum Performance requirements for RATE_REDUC operating points 4 and 7 of SO 70 (which exactly correspond to the RATE_REDUC operating points 0 and 7 of SO 68).

Therefore, the main parameters in the decision tree are:

a) 16 kHz support in the implementation, and

b) SO 68 compliance of the test implementation.

Depending on the implementation profile of the device under test, one of 4 possible Test Suites are to be used to demonstrate SO 70 compliance. These 4 test suites named Test suites A, B, C, D, and the individual tests comprising the Test suites are highlighted in Table 2.3.2-1.

Table 2.3.2-1 Test Suites for SO 70 compliance

Test Suites	Set of Experiments	Notes
A	Experiment 1, 2 and 7	Mobile/MGW already supporting SO 68 compliance
B	Experiment 1, 2, 3, 4, 7 and 8	Mobile/MGW NOT already supporting SO 68 compliance
C	Experiment 5, 6, and 8	Infra/MGW already supporting SO 68 compliance
D	Experiment 3, 4, and 8	Infra/MGW NOT already supporting SO 68 compliance

Each of the individual experiments are further defined in detail by Table 2.3.2-2.

Table 2.3.2-2 Experiments for SO 70 compliance

Experiment	Individual tests	Notes
1	WB clean/level/FER/signaling - ACR	Mobile supporting 16 kHz Fs
2	WB noise/FER - P.835	Mobile supporting 16 kHz Fs
3	NB clean/level/FER/signaling including SO 68 interoperable mode tests - ACR	BS supporting 8 kHz, and MS supporting 8/16 kHz) - SO 68 compliance not PROVEN
4	NB noise/FER including SO 68 interoperable mode tests - P.835	BS supporting 8 kHz, and MS supporting 8/16 kHz) - SO 68 compliance not PROVEN
5	NB clean/level/FER/signaling NOT including SO 68 interoperable mode tests - ACR	BS supporting 8 kHz - SO 68 compliance already PROVEN
6	NB noise/FER/signaling NOT including SO 68 interoperable mode tests - P.835	BS supporting 8 kHz - SO 68 compliance already PROVEN
7	WB music decoder test – ACR	Mobile supporting 16 kHz Fs
8	NB music decoder test – ACR	BS supporting 8 kHz Fs

2.3.2.1 Definition

The codec subjective test is intended to validate the implementation of the speech codec being tested using the master codec defined in 3.3.3 as a reference. Experiments 1, 3 and 5 are based on the Absolute Category Rating (ACR) method, which yields the Mean Opinion Score (MOS) as described in [10]. Experiments 2, 4 and 6 are based on the ITU-T Recommendation P.835 described in [13].

2.3.2.2 Method of Measurement

The subjective tests involve a listening-only assessment of the quality of the codec being tested, using the master codec as a reference. Subjects from the general population of telephone users will rate the various conditions of the test. Material supplied with this standard for use with this test includes source speech, impaired packet files from the master codec encoder, and source speech processed by various Modulated Noise Reference Unit (MNRU) conditions and other references. The basic Absolute Category Rating test procedure involves rating all conditions using a five-point scale describing the opinion of the test condition. This procedure is fully described in [10]. The P.835 test method involves rating all conditions on scales of “Signal”, “Background”, and “Overall” quality and is fully described in [13].

2.3.2.3 Test Conditions and Test Design for SO 70

Listening experiments 1, 3, 5 for SO 70 are performed as ACR listening tests. Experiments 2, 4, and 6 for SO 70 are performed as P.835 listening tests.

2.3.2.3.1 Subjective Experiment 1 for SO 70

The Test Parameters for Listening Experiment 1 are presented in Table 2.3.2.3.1-1.

Condition	Description
Type of test	MOS (P.800), Wideband
Number of talkers	4 males, 4 females
Background noise	none (ambient)
Audio Input Level	-22 dB, -32 dB, -12 dB
Filter characteristics	P.341 (refer Section 3.3.2.4)
Reference conditions	(8) Specified reference conditions
Test conditions	o Low Audio Input Level -32 dB + 1% d&b o Nominal Audio Input Level, -22 dB o High Audio Input Level -12 dB o 3% FER and 1%FER + 2%pls at Nominal Audio Input Level, -22
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

Table 2.3.2.3.1-1 SO 70 Listening Experiment 1 Test Parameters

Condition	Description
Type of test	MOS (P.800), Wideband
Number of talkers	4 males, 4 females
Background noise	none (ambient)
Audio Input Level	-22 dB, -32 dB, -12 dB
Filter characteristics	P.341 (refer Section 3.3.2.4)
Reference conditions	(8) Specified reference conditions
Test conditions	o Low Audio Input Level -32 dB + 1% d&b o Nominal Audio Input Level, -22 dB o High Audio Input Level -12 dB

Condition	Description
	○ 3% FER and 1%FER + 2%pls at Nominal Audio Input Level, -22
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

The Test Conditions for Listening Experiment 1 are presented in Table 2.3.2.3.1-2.

Table 2.3.2.3.1-2 SO 70 Listening Experiment 1 Test Conditions

Exp.1	Wideband - ACR	
Reference Conditions		
File	**MNRU**	
a01	7dB MNRU	Reference
a02	14dB MNRU	Reference
a03	21dB MNRU	Reference
a04	28dB MNRU	Reference
a05	35dB MNRU	Reference
a06	42dB MNRU	Reference
a07	49dB MNRU	Reference
a08	Direct Source	Reference
Test Conditions		
File	**Condition**	**Enc-Dec**
a09	Nominal level	M-M
a10	Nominal level	M-T
a11	Nominal level	T-T
a12	Nominal level	T-M
a13	Low level, 1% d&b	M-M
a14	Low level, 1% d&b	M-T
a15	Low level, 1% d&b	T-T
a16	Low level, 1% d&b	T-M
a17	High level	M-M
a18	High level	M-T
a19	High level	T-T
a20	High level	T-M
a21	1% FER, 1% PLS	M-M
a22	1% FER, 1% PLS	M-T
a23	3% FER	M-M
a24	3% FER	M-T

1

2 2.3.2.3.2 Subjective Experiment 2 for SO 70

3 The Test Parameters for Listening Experiment 2 are presented in Table 2.3.2.3.2-1.

4 **Table 2.3.2.3.2-1 SO 70 Listening Experiment 2 Test Parameters**

Condition	Description
Type of test	P-NSA (P.835), Wideband
Number of talkers	3 males, 3 females
Background noise	Specified test conditions
Audio Input Level	-22 dB
Filter characteristics	P.341 (refer Section 3.3.2.4)
Reference conditions	(8) Specified reference conditions
Test conditions	o Car Noise, 10 dB SNR o Car Noise, 20 dB SNR + 2% FER o Street Noise, 15 dB SNR o Babble noise, 20 dB S/N
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

5

Condition	**Description**
Type of test	P-NSA (P.835), Wideband
Number of talkers	3 males, 3 females
Background noise	Specified test conditions
Audio Input Level	-22 dB
Filter characteristics	P.341 (refer Section 3.3.2.4)
Reference conditions	(8) Specified reference conditions
Test conditions	o Car Noise, 10 dB SNR o Car Noise, 20 dB SNR + 2% FER o Street Noise, 15 dB SNR o Babble noise, 20 dB S/N
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

6

7 The Test Conditions for Listening Experiment 2 are presented in Table 2.3.2.3.2-2

Table 2.3.2.3.2-2 SO 70 Listening Experiment 2 Test Conditions.

Exp.2	Wideband - P.835	
Reference Conditions		
File	**MNRU, SNR**	
b01	MNRU=40dB, SNR=40dB	Reference
b02	MNRU=40dB, SNR=20dB	Reference
b03	MNRU=40dB, SNR=0dB	Reference
b04	MNRU=0dB, SNR=40dB	Reference
b05	MNRU=20dB, SNR=40dB	Reference
b06	MNRU=10dB, SNR=10dB	Reference
b07	MNRU=20dB, SNR=20dB	Reference
b08	MNRU=40dB, SNR=30dB	Reference
Test Conditions		
File	**Condition**	**Enc-Dec**
b09	Car 10dB SNR	M-M
b10	Car 10dB SNR	M-T
b11	Car 10dB SNR	T-T
b12	Car 10dB SNR	T-M
b13	Car 20dB SNR + 2% FER	M-M
b14	Car 20dB SNR + 2% FER	M-T
b15	Car 20dB SNR + 2% FER	T-T
b16	Car 20dB SNR + 2% FER	T-M
b17	Street 15dB SNR	M-M
b18	Street 15dB SNR	M-T
b19	Street 15dB SNR	T-T
b20	Street 15dB SNR	T-M
b21	Babble 20dB SNR	M-M
b22	Babble 20dB SNR	M-T
b23	Babble 20dB SNR	T-T
b24	Babble 20dB SNR	T-M

2.3.2.3.3 Subjective Experiment 3 for SO 70

The Test Parameters for Listening Experiment 3 are presented in Table 2.3.2.3.3-1..

Table 2.3.2.3.3-1 SO 70 Listening Experiment 3 Test Parameters

Condition	Description
Type of test	ACR (P.800), Narrowband
Number of talkers	4 males, 4 females
Background noise	none (ambient)
Audio Input Level	-22 dB, -32 dB, -12 dB
Filter characteristics	MIRS
Reference conditions	(8) Specified reference conditions
Test conditions	o Nominal level, Modes 0, 4, 7 o Low level, Modes 0, 4 o High Level, Mode 0, 4 o 1% d&b, 1% pls, Modes 0, 4 o 3% FER, Modes 0, 4
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

The Test Conditions for Listening Experiment 3 are presented in Table 2.3.2.3.3-2

Table 2.3.2.3.3-2 SO 70 Listening Experiment 3 Test Conditions.

Exp.3	Narrowband - ACR		
Reference Conditions			
File	**MNRU**		
c01	5dB MNRU		Reference
c02	10dB MNRU		Reference
c03	15dB MNRU		Reference
c04	20dB MNRU		Reference
c05	25dB MNRU		Reference
c06	30dB MNRU		Reference
c07	35dB MNRU		Reference
c08	Direct Source		Reference
Test Conditions			
File	**Condition**		**Enc-Dec**
c09	Nominal, Mode 0	LB portion of Wideband mode - decoder test only	M-M
c10	Nominal, Mode 0	LB portion of Wideband mode - decoder test only	M-T
c11	Nominal, Mode 4	(interoperable with Mode 0 of SO 68) support	M-M
c12	Nominal, Mode 4	(interoperable with Mode 0 of SO 68) support	M-T
c13	Nominal, Mode 4	(interoperable with Mode 0 of SO 68) support	T-T

Exp.3	Narrowband - ACR		
c14	Nominal, Mode 4	(interoperable with Mode 0 of SO 68) support	T-M
c15	Nominal, Mode 7	(interoperable with Mode 0 of SO 68) support	T-T
c16	Nominal, Mode 7	(interoperable with Mode 0 of SO 68) support	T-M
c17	Nominal, Mode 7	(interoperable with Mode 0 of SO 68) support	M-M
c18	Nominal, Mode 7	(interoperable with Mode 0 of SO 68) support	M-T
c19	Low, Mode 0	LB portion of Wideband mode - decoder test only	M-M
c20	Low, Mode 0	LB portion of Wideband mode - decoder test only	M-T
c21	Low, Mode 4	(interoperable with Mode 0 of SO 68) support	M-M
c22	Low, Mode 4	(interoperable with Mode 0 of SO 68) support	M-T
c23	Low, Mode 4	(interoperable with Mode 0 of SO 68) support	T-T
c24	Low, Mode 4	(interoperable with Mode 0 of SO 68) support	T-M
c25	High, Mode 0	LB portion of Wideband mode - decoder test only	M-M
c26	High, Mode 0	LB portion of Wideband mode - decoder test only	M-T
c27	High, Mode 4	(interoperable with Mode 0 of SO 68) support	M-M
c28	High, Mode 4	(interoperable with Mode 0 of SO 68) support	M-T
c29	High, Mode 4	(interoperable with Mode 0 of SO 68) support	T-T
c30	High, Mode 4	(interoperable with Mode 0 of SO 68) support	T-M
c31	Mode 0, 1% D&B, 1% PLS	LB portion of Wideband mode - decoder test only	M-M
c32	Mode 0, 1% D&B, 1% PLS	LB portion of Wideband mode - decoder test only	M-T
c33	Mode 4, 1% D&B, 1% PLS	(interoperable with Mode 0 of SO 68) support	M-M
c34	Mode 4, 1% D&B, 1% PLS	(interoperable with Mode 0 of SO 68) support	M-T
c35	Mode 4, 1% D&B, 1% PLS	(interoperable with Mode 0 of SO 68) support	T-T
c36	Mode 4, 1% D&B, 1% PLS	(interoperable with Mode 0 of SO 68) support	T-M
c37	Mode 0, 3% FER	LB portion of Wideband mode - decoder test only	M-M
c38	Mode 0, 3% FER	LB portion of Wideband mode - decoder test only	M-T
c39	Mode 4, 3% FER	(interoperable with Mode 0 of SO 68) support	M-M
c40	Mode 4, 3% FER	(interoperable with Mode 0 of SO 68) support	M-T

2.3.2.3.4 Subjective Experiment 4 for SO 70

The Test Parameters for Listening Experiment 4 are presented in Table 2.3.2.3.4-1.

Table 2.3.2.3.4-1 SO 70 Listening Experiment 4 Test Parameters

Condition	Description
Type of test	P-NSA (P.835), Narrowband
Number of talkers	3 males, 3 females
Background noise	Specified test conditions
Audio Input Level	-22 dB
Filter characteristics	MIRS
Reference conditions	(8) Specified reference conditions
Test conditions	o Car Noise, 15 dB SNR, Modes 0, 4, 7 o Street Noise, 15 dB SNR, Modes 0, 4 o Babble noise, 20 dB SNR, 2%FER, Modes 0, 4
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

The Test Conditions for Listening Experiment 4 are presented in Table 2.3.2.3.4-2

Table 2.3.2.3.4-2 SO 70 Listening Experiment 4 Test Conditions.

Exp.4	Narrowband - P.835		
Reference Conditions			
File	**MNRU**		
d01	MNRU=40dB, SNR=40dB		Reference
d02	MNRU=40dB, SNR=20dB		Reference
d03	MNRU=40dB, SNR=0dB		Reference
d04	MNRU=0dB, SNR=40dB		Reference
d05	MNRU=20dB, SNR=40dB		Reference
d06	MNRU=10dB, SNR=10dB		Reference
d07	MNRU=20dB, SNR=20dB		Reference
d08	MNRU=40dB, SNR=30dB		Reference
Test Conditions			
File	**Condition**		**Enc-Dec**
d09	Car 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-M
d10	Car 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-T
d11	Car 15dB SNR, Mode 4	(interoperable with Mode 0 of SO 68) support	M-M
d12	Car 15dB SNR, Mode 4	(interoperable with Mode 0 of SO 68) support	M-T
d13	Car 15dB SNR, Mode 4	(interoperable with Mode 0 of SO 68) support	T-T
d14	Car 15dB SNR, Mode 4	(interoperable with Mode 0 of SO 68) support	T-M
d15	Car 15dB SNR, Mode 7	(interoperable with Mode 0 of SO 68) support	T-T
d16	Car 15dB SNR, Mode 7	(interoperable with Mode 0 of SO 68) support	T-M

Exp.4	Narrowband - P.835		
d17	Car 15dB SNR, Mode 7	(interoperable with Mode 0 of SO 68) support	M-M
d18	Car 15dB SNR, Mode 7	(interoperable with Mode 0 of SO 68) support	M-T
d19	Street 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-M
d20	Street 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-T
d21	Street 15dB SNR, Mode 4	(interoperable with Mode 0 of SO 68) support	M-M
d22	Street 15dB SNR, Mode 4	(interoperable with Mode 0 of SO 68) support	M-T
d23	Street 15dB SNR, Mode 4	(interoperable with Mode 0 of SO 68) support	T-T
d24	Street 15dB SNR, Mode 4	(interoperable with Mode 0 of SO 68) support	T-M
d25	Babble 20dB SNR, 2% FER, Mode 0	LB portion of Wideband mode - decoder test only	M-M
d26	Babble 20dB SNR, 2% FER, Mode 0	LB portion of Wideband mode - decoder test only	M-T
d27	Babble 20dB SNR, 2% FER, Mode 4	(interoperable with Mode 0 of SO 68) support	M-M
d28	Babble 20dB SNR, 2% FER, Mode 4	(interoperable with Mode 0 of SO 68) support	M-T
d29	Babble 20dB SNR, 2% FER, Mode 4	(interoperable with Mode 0 of SO 68) support	T-T
d30	Babble 20dB SNR, 2% FER, Mode 4	(interoperable with Mode 0 of SO 68) support	T-M

2.3.2.3.5 Subjective Experiment 5 for SO 70

The Test Parameters for Listening Experiment 5 are presented in Table 2.3.2.3.5-1.

Table 2.3.2.3.5-1 SO 70 Listening Experiment 5 Test Parameters

Condition	Description
Type of test	ACR (P.800), Narrowband
Number of talkers	4 males, 4 females
Background noise	none (ambient)
Audio Input Level	-22 dB, -32 dB, -12 dB
Filter characteristics	MIRS
Reference conditions	(8) Specified reference conditions
Test conditions	o Nominal level, Mode 0 o Low level, Mode 0 o High level, Mode 0 o Nominal level, Mode 0, 1% d&b o Nominal level, Mode 0, 10% d&b o 2% FER, Mode 0, 1% d&b o 6% FER, Mode 0, 10% d&b o Nominal, Mode 0, 1% pls
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

The Test Conditions for Listening Experiment 5 are presented in Table 2.3.2.3.5-2

Table 2.3.2.3.5-2 SO 70 Listening Experiment 5 Test Conditions.

Exp.5	Narrowband - ACR		
Reference Conditions			
File	**MNRU**		
e01	5dB MNRU		Reference
e02	10dB MNRU		Reference
e03	15dB MNRU		Reference
e04	20dB MNRU		Reference
e05	25dB MNRU		Reference
e06	30dB MNRU		Reference
e07	35dB MNRU		Reference
e08	Direct Source		Reference
Test Conditions			
File	**Condition**		**Enc-Dec**
e09	Nominal, Mode 0	LB portion of Wideband mode - decoder test only	M-M
e10	Nominal, Mode 0	LB portion of Wideband mode - decoder test only	M-T
e11	Low, Mode 0	LB portion of Wideband mode - decoder test only	M-M
e12	Low, Mode 0	LB portion of Wideband mode - decoder test only	M-T
e13	High, Mode 0	LB portion of Wideband mode - decoder test only	M-M
e14	High, Mode 0	LB portion of Wideband mode - decoder test only	M-T
e15	Nominal, Mode 0, 1% D&BS	LB portion of Wideband mode - decoder test only	M-M
e16	Nominal, Mode 0, 1% D&BS	LB portion of Wideband mode - decoder test only	M-T
e17	Nominal, Mode 0, 10% D&BS	(interoperable with Mode 0 of SO 68) support	M-M
e18	Nominal, Mode 0, 10% D%BS	(interoperable with Mode 0 of SO 68) support	M-T
e19	FER 2%, Mode 0, 1% D&BS	LB portion of Wideband mode - decoder test only	M-M
e20	FER 2%, Mode 0, 1% D&BS	LB portion of Wideband mode - decoder test only	M-T
e21	FER 6%, Mode 0, 10% D&BS	(interoperable with Mode 0 of SO 68) support	M-M
e22	FER 6%, Mode 0, 10% D%BS	(interoperable with Mode 0 of SO 68) support	M-T
e23	Nominal, Mode 0, 1% PLS	LB portion of Wideband mode - decoder test only	M-M
e24	Nominal, Mode 0, 1% PLS	LB portion of Wideband mode - decoder test only	M-T

2.3.2.3.6 Subjective Experiment 6 for SO 70

The Test Parameters for Listening Experiment 6 are presented in Table 2.3.2.3.6-1.

Table 2.3.2.3.6-1 SO 70 Listening Experiment 6 Test Parameters

Condition	Description
Type of test	P-NSA (P.835),Narrowband
Number of talkers	3 males, 3 females
Background noise	Specified test conditions
Audio Input Level	-22 dB
Filter characteristics	MIRS
Reference conditions	(8) Specified reference conditions
Test conditions	o Car Noise, 15 dB SNR, Mode 0 o Street Noise, 15 dB SNR, Mode 0 o Babble, 20 dB SNR, 2%FER, Mode 0 o Car Noise, 15 dB SNR, Mode 0, 2% d&b o Car Noise, 15 dB SNR, Mode 0, 1% pls
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

The Test Conditions for Listening Experiment 6 are presented in Table 2.3.2.3.6-2

Table 2.3.2.3.6-2 SO 70 Listening Experiment 6 Test Conditions.

Exp.6	**Narrowband - P.835**		
Reference Conditions			
File	**MNRU**		
f01	MNRU=40dB, SNR=40dB		Reference
f02	MNRU=40dB, SNR=20dB		Reference
f03	MNRU=40dB, SNR=0dB		Reference
f04	MNRU=0dB, SNR=40dB		Reference
f05	MNRU=20dB, SNR=40dB		Reference
f06	MNRU=10dB, SNR=10dB		Reference
f07	MNRU=20dB, SNR=20dB		Reference
f08	MNRU=40dB, SNR=30dB		Reference
Test Conditions			
File	**Condition**		**Enc-Dec**
f09	Car 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-M
f10	Car 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-T
f11	Street 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-M
f12	Street 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-T
f13	Babble 20dB SNR, 2% FER, Mode 0	LB portion of Wideband mode - decoder test only	M-M

Exp.6	Narrowband - P.835		
f14	Babble 20dB SNR, 2% FER, Mode 0	LB portion of Wideband mode - decoder test only	M-T
f15	Car 20dB SNR, 2% d&b, Mode 0	LB portion of Wideband mode - decoder test only	M-M
f16	Car 20dB SNR, 2% d&b, Mode 0	LB portion of Wideband mode - decoder test only	M-T
f15	Car 20dB SNR, 1% pls, Mode 0	LB portion of Wideband mode - decoder test only	M-M
f16	Car 20dB SNR, 1% pls, Mode 0	LB portion of Wideband mode - decoder test only	M-T

1

2 2.3.2.3.7 Subjective Experiment 7 for SO 70

3 The Test Parameters for Listening Experiment 7 are presented in Table 2.3.2.3.7-1.

4 **Table 2.3.2.3.7-1 SO 70 Listening Experiment 7 Test Parameters**

Condition	Description
Type of test	ACR (P.800), Wideband
Number of genres	4
Background noise	none (ambient)
Audio Input Level	-22 dB
Filter characteristics	P.341 (refer Section 3.3.2.4)
Reference conditions	(4) Specified reference conditions
Test conditions	0% FER and 3% FER
Encoder/Decoder Combinations	(2) M/M, M/T

5

6 The Test Conditions for Listening Experiment 7 are presented in Table 2.3.2.3.7-2.

7 **Table 2.3.2.3.7-2 SO 70 Listening Experiment 7 Test Conditions.**

Exp.7	Wideband Music	
File	**Reference Condition**	
g01	MNRU=15dB	Reference
g02	MNRU=25dB	Reference
g03	MNRU=35dB	Reference
g04	Source	Reference
File	**Test Condition**	**Enc-Dec**
g05	0% FER	M-M
g06	0% FER	M-T
g07	3% FER	M-M
g08	3% FER	M-T

2.3.2.3.8 Subjective Experiment 8 for SO 70

The Test Parameters for Listening Experiment 8 are presented in Table 2.3.2.3.8-1.

Table 2.3.2.3.8-1 SO 70 Listening Experiment 8 Test Parameters

Condition	Description
Type of test	ACR (P.800), Narrowband
Number of genres	4
Background noise	none (ambient)
Audio Input Level	-22 dB
Filter characteristics	MIRS
Reference conditions	(4) Specified reference conditions
Test conditions	0% FER and 3% FER
Encoder/Decoder Combinations	(2) M/M, M/T

The Test Conditions for Listening Experiment 8 are presented in Table 2.3.2.3.8-2.

Table 2.3.2.3.8-2 SO 70 Listening Experiment 8 Test Conditions.

Exp.8	Narrowband Music	
File	**Reference Condition**	
h01	MNRU=10dB	Reference
h02	MNRU=20dB	Reference
h03	MNRU=30dB	Reference
h04	Source	Reference
File	**Test Condition**	**Enc-Dec**
h05	0% FER	M-M
h06	0% FER	M-T
h07	3% FER	M-M
h08	3% FER	M-T

2.3.2.3.9 Numerical Parameters for the SO 70 Listening Experiments

Table 2.3.2.3.9-1 describes the resultant numerology that is used for the eight SO 70 listening experiments. The first column is the description of the parameter and columns 2 to 9 show the numerical value for each of the parameters, for the eight listening experiments. For each listening experiment, the different Encode/Decode Test conditions include various interconnections between the Master and Test Encoders, and the Master and Test Decoders. There are eight reference

conditions in each of the experiments1 through 6, and four reference conditions in experiments 7 and 8.

Table 2.3.2.3.9-1 Numerical Parameters for the SO 70 Listening Experiments

Parameter	Exp.1	Exp.2	Exp.3	Exp.4	Exp.5	Exp.6	Exp.7	Exp.8
Type of test	ACR	P.835	ACR	P.835	ACR	P.835	ACR	ACR
Encode/Decode Test conditions	16	16	32	22	16	8	4	4
Reference Conditions	8	8	8	8	8	8	4	4
Total Conditions	24	24	40	30	24	16	8	8
Talkers (* genres)	8	6	8	6	8	6	3*	3*
Stimuli per Talker (* genres)	8	8	8	8	8	8	4*	4*
Stimuli per Condition	64	48	64	48	64	48	12	12
Total Stimuli per Experiment	1536	1152	2560	1440	1536	768	96	96
Listening Panels	8	8	8	8	8	8	4	4
Stimuli per Listening Panel	192	144	320	180	192	96	24	24
Listeners (Voters)	32	32	32	32	32	32	32	32
Listeners (Voters) per Listening Panel	4	4	4	4	4	4	8	8
Votes per Condition	256	192	256	192	256	192	96	96

2.3.3 Speech Material for SO 70 Testing

The source speech files used for SO 70 compliance testing consist of Harvard sentences pairs, which are preprocessed to include proper level adjustment and noise mixing for use in the subjective experiments. The talkers used in these files consist of adult males and adult females, and are native speakers of North American English.

For the following discussion, it may be useful to refer to Table 4-3 for the composition of the Software Distribution database.

The source speech material for subjective Experiments is contained in directory /so70/subjctv/exp*/source. Each file has been appropriately pre-filtered, level adjusted, and noise-processed. These files are named src.*. The speech database also includes samples processed through the various reference conditions in directory /so70/subjctv/exp*/ref. The reference conditions are named ref.* for the respective conditions given in the tables in Section 2.3.2.3.

2.3.4 Processing of Speech Material for SO 70 Testing

The source speech material shall be processed by the various combinations of encoders and decoders listed in the descriptions of the experiments given in Section 2.3.2. The master codec software described in Section 3.3.3 shall be used in the processing involving the master codec. Generally, the master codec encoder and decoder outputs have been provided in the respective

directories, /so70/subjctv/exp*/m_pkt and /so70/subjctv/exp*/m_m. Execution of the master codec software is needed only for the test encoder/master decoder combination for each experiment/condition.

All codec processing shall be done digitally. Noise suppression and post-filter options shall be enabled for both the master and the test codecs. The digital format of the speech files is described in Section 3.3.4.4.

The naming convention of the processed speech is as follows: For the packet files in the /so70/subjctv/exp{1,3,5}/m_pkt directory, the *.p12 files are the master packet files for the *.s12 source file. Likewise, the *.p22 and *.p32 files are the respective packet files for the *.s22 and *.s32 source files. The *.pf3 files are the impaired packet files which will be described in Section 2.3.4.3.

Similarly, the directory /so70/subjctv/exp{2,4,6}/m_pkt contains the master packet files for the respective experiments. Here, the *.pc10, *.pb20, and *.ps files are the master packet files for the *.c15, *.b20, and *.s15 source files, respectively.

For the master encode/master decode directories (/so70/subjctv/exp*/m_m), the naming convention of the speech files is such that the first two characters of the file name indicate the codec combination and the suffix indicates the condition numbers in Table 2.3.2.3.1-2 and Table 2.3.2.3.2-2

Naming conventions for the remaining two experiments follow accordingly.

Detailed descriptions of all processing operations are given in Section 6 .

2.3.4.1 Encoding by the Test Codec

All of the source files will be encoded by the test codec to produce encoded packet files. For ease of reference, it is recommended that directories /so70/subjctv/exp*/t_pkt be created to deposit the test encoder output packets, and that the naming conventions be made consistent with the master codec.

2.3.4.2 Decoding by the Master/Test Codecs

The encoded packet files generated from the various encoders/conditions shall be processed through the master and test decoders.

2.3.4.3 Introduction of Impairments

For the frame error conditions, the impaired master codec encoded packet files are provided in the /so70/subjctv/exp*/m_pkt directory. Unlike other conditions, this condition uses only the test decoder and not the test encoder.

For the Dim-and-Burst processing, and also the Packet Level Signaling conditions, the processing requires inputs from a signaling file to control maximum encoding rate. An external software utility (Evrc_wb_iwf in Section 3.3.2.3) is also needed to reduce the data rate of certain packets from full rate to half rate. Details of these operations are given in Section 6. The signaling file and other utilities are provided in /so70/tools/ directory.

2.3.4.4 Ensuring Proper Encoded Frame Packet Files

All encoded frame packet files shall be examined to ensure that the files only contain data in those file locations where data should exist for a given data rate.

The examination of the encoded frame packet files should indicate the occurrence of any improper data in the files but the examination must not alter the encoded frame packet files in any way.

2.3.4.5 Post-processing of test-condition output files

In order to build the play sets to be presented to the listening panels, the output files for the various test conditions must be processed to provide the appropriate listening conditions. In addition, the concatenated output files must be partitioned into the samples representing the combination of test-condition and talker. The listening conditions for Narrowband experiments are provided by filtering the output files using the STL software tool (*filter*) with the MIRS-receive filter mask. The listening conditions for Wideband experiments are provided by mixing (STL tool *oper*) the output files with Psophometrically filtered noise (STL tool *filter*, PSO filter mask) at 74dBov. STL tool *astrip* is also used to split the concatenated files into the individual samples appropriate for the experiment. Table 2.3.4.5-1 shows the cutting-points to be used with the *astrip* tool for producing the two-sentence samples for the Experiments 1, 3, and 5 ACR test. Table 2.3.4.5-2 shows the cutting-points to be used with the *astrip* tool for producing the single-sentence sub-samples for the Experiments 2, 4, and 6 P.835 test. Table 2.3.4.5-3 shows the sub-samples that make up the samples (i.e., sentence triads) for the P.835 test.

Table 2.3.4.5-1 Cutting Points for the astrip Software Tool for the SO 70 Experiments 1, 3, and 5 ACR Test

Experiment I - ACR							
Sentence-pair		Start sample	Length (samples)	Sentence-pair		Start sample	Length (samples)
1	m1p1	1	113706	33	m1p5	3615102	113446
2	f1p1	113707	118586	34	f1p5	3728548	114249
3	m2p1	232293	111900	35	m2p5	3842797	115062
4	f2p1	344193	117486	36	f2p5	3957859	117344
5	m3p1	461679	110993	37	m3p5	4075203	129258
6	f3p1	572672	123570	38	f3p5	4204461	117851
7	m4p1	696242	106749	39	m4p5	4322312	105606
8	f4p1	802991	110876	40	f4p5	4427918	111339
9	m1p2	913867	102934	41	m1p6	4539257	111723
10	f1p2	1016801	120088	42	f1p6	4650980	105818
11	m2p2	1136889	124661	43	m2p6	4756798	109458
12	f2p2	1261550	111121	44	f2p6	4866256	122664
13	m3p2	1372671	110603	45	m3p6	4988920	116439
14	f3p2	1483274	126079	46	f3p6	5105359	127468
15	m4p2	1609353	99074	47	m4p6	5232827	109566
16	f4p2	1708427	108801	48	f4p6	5342393	108807
17	m1p3	1817228	102960	49	m1p7	5451200	118850
18	f1p3	1920188	118392	50	f1p7	5570050	111097
19	m2p3	2038580	121905	51	m2p7	5681147	121218
20	f2p3	2160485	120916	52	f2p7	5802365	116957
21	m3p3	2281401	104536	53	m3p7	5919322	112149
22	f3p3	2385937	108073	54	f3p7	6031471	110715
23	m4p3	2494010	97510	55	m4p7	6142186	101539
24	f4p3	2591520	107375	56	f4p7	6243725	118876
25	m1p4	2698895	116301	57	m1p8	6362601	118866
26	f1p4	2815196	105011	58	f1p8	6481467	116617
27	m2p4	2920207	124711	59	m2p8	6598084	130938
28	f2p4	3044918	115760	60	f2p8	6729022	123975
29	m3p4	3160678	119447	61	m3p8	6852997	115326
30	f3p4	3280125	110386	62	f3p8	6968323	121531
31	m4p4	3390511	111967	63	m4p8	7089854	104458
32	f4p4	3502478	112624	64	f4p8	7194312	102903

Table 2.3.4.5-2 Cutting Points for the *astrip* Software Tool for the SO 70 Experiments 2, 4, and 6 P.835 Test

Experiment II - P.835											
Sentence		Start sample	Length (samples)	Sentence		Start sample	Length (samples)	Sentence		Start sample	Length (samples)
1	m1s01	1	57758	33	m3s05	1853425	56706	65	m2s11	3679398	55140
2	m1s02	57759	54087	34	m3s06	1910131	49310	66	m2s12	3734538	55821
3	f1s01	111846	58737	35	f3s05	1959441	55546	67	f2s11	3790359	61656
4	f1s02	170583	60083	36	f3s06	2014987	50615	68	f2s12	3852015	60100
5	m2s01	230666	57758	37	m1s07	2065602	62653	69	m3s11	3912115	55432
6	m2s02	288424	56412	38	m1s08	2128255	54048	70	m3s12	3967547	59225
7	f2s01	344836	59226	39	f1s07	2182303	50470	71	f3s11	4026772	64087
8	f2s02	404062	59226	40	f1s08	2232773	54338	72	f3s12	4090859	63406
9	m3s01	463288	51884	41	m2s07	2287111	60526	73	m1s13	4154265	62290
10	m3s02	515172	59593	42	m2s08	2347637	62846	74	m1s14	4216555	55664
11	f3s01	574765	65834	43	f2s07	2410483	56078	75	f1s13	4272219	56276
12	f3s02	640599	54943	44	f2s08	2466561	62846	76	f1s14	4328495	55460
13	m1s03	695542	47316	45	m3s07	2529407	60236	77	m2s13	4383955	64737
14	m1s04	742858	57323	46	m3s08	2589643	55208	78	m2s14	4448692	57193
15	f1s03	800181	58954	47	f3s07	2644851	56465	79	f2s13	4505885	61169
16	f1s04	859135	59389	48	f3s08	2701316	55691	80	f2s14	4567054	55052
17	m2s03	918524	64366	49	m1s09	2757007	57238	81	m3s13	4622106	54440
18	m2s04	982890	58954	50	m1s10	2814245	55305	82	m3s14	4676546	55970
19	f2s03	1041844	54821	51	f1s09	2869550	62506	83	f3s13	4732516	58213
20	f2s04	1096665	60042	52	f1s10	2932056	53308	84	f3s14	4790729	53013
21	m3s03	1156707	51558	53	m2s09	2985364	61148	85	m1s15	4843742	56582
22	m3s04	1208265	57758	54	m2s10	3046512	54353	86	m1s16	4900324	61688
23	f3s03	1266023	64393	55	f2s09	3100865	59266	87	f1s15	4962012	59292
24	f3s04	1330416	57758	56	f2s10	3160131	61148	88	f1s16	5021304	59083
25	m1s05	1388174	59172	57	m3s09	3221279	59580	89	m2s15	5080387	65752
26	m1s06	1447346	47569	58	m3s10	3280859	64388	90	m2s16	5146139	63251
27	f1s05	1494915	55981	59	f3s09	3345247	59371	91	f2s15	5209390	64398
28	f1s06	1550896	60912	60	f3s10	3404618	58012	92	f2s16	5273788	60646
29	m2s05	1611808	61492	61	m1s11	3462630	54265	93	m3s15	5334434	54602
30	m2s06	1673300	61492	62	m1s12	3516895	56113	94	m3s16	5389036	60125
31	f2s05	1734792	57721	63	f1s11	3573008	54459	95	f3s15	5449161	59083
32	f2s06	1792513	60912	64	f1s12	3627467	51931	96	f3s16	5508244	61792

Table 2.3.4.5-3 Composition of the Sentence-Triad Samples for the Experiments 2, 4, and 6 P.835 Test

Sentence-triad	Sentence 1	Sentence 2	Sentence 3
t1	s01	s02	s03
t2	s04	s05	s06
t3	s07	s08	s09
t4	s10	s11	s12
t5	s13	s14	s15
t6	s16	s01	s02
t7	s03	s04	s05
t8	s06	s07	s08

2.3.5 Randomization

For each of the first six subjective experiments, each presentation sample consists of a speech sample processed under a condition of the test. For the ACR Experiments 1, 3, and 5 the sample consists of a pair of concatenated sentences of approximately 8 sec. duration. For the P.835 Experiments 2, 4, and 6 the sample consists of three sub-samples, where each sub-sample is a single sentence of approximately 4 sec. duration. The samples shall be presented to the listeners in a

randomized presentation order. The listeners for each file set shall be presented with practice trials for subjective Experiments 1, 3, and 5 and Experiments 2, 4, and 6. The randomization of the test samples has been accomplished with the following constraints for each of the six experiments:

1. A trial, i.e., a test sample, for the combination of each test condition and each talker shall be presented exactly once to each listening panel (i.e., # trials/panel = # conditions x # talkers).
2. Randomization is in “blocks”, such that one sample of each test condition is presented once, with a randomly selected talker, in each block. This ensures that listeners rate each test condition equally often in the initial, middle and final parts of the block and controls for the effects of time and order of presentation. A block contains the same number of samples as there are test-conditions involved in the test. A test “session” consists of the same number of blocks as there are talkers involved in the test. Each session is presented to a listening panel of four listeners.
3. Randomizations are constructed such that talker gender is alternated on successive trials resulting in the same talker never being presented on consecutive trials.

Table 2.3.5-1 shows an example randomization for a single listening panel. Each entry in the table is the file name for a sample with the following file-naming convention - ***xxyy.zzz***, where ***xx*** is the talker, ***yy*** is the sample, and ***zzz*** is the test condition.

Table 2.3.5-1 Example Randomization for the Experiments 1, 3, and 5 ACR Test

Panel 1	Blk 1	Blk 2	Blk 3	Blk 4	Blk 5	Blk 6	Blk 7	Blk 8
1	a12f1s6	a18f4s8	a03f2s1	a11f3s6	a06f2s6	a01f2s5	a05f3s8	a09f4s8
2	a11m4s5	a04m2s1	a08m3s7	a22m2s6	a04m4s4	a06m3s5	a15m1s8	a19m2s7
3	a20f1s6	a16f2s1	a21f4s8	a10f2s1	a14f2s2	a03f4s7	a12f2s2	a15f2s4
4	a16m1s1	a20m2s5	a23m2s3	a06m2s4	a17m1s8	a23m4s2	a01m3s3	a20m3s1
5	a05f2s4	a08f2s4	a05f4s8	a19f3s6	a05f1s6	a09f2s4	a13f3s3	a14f1s4
6	a18m3s1	a19m1s1	a06m1s7	a07m3s1	a01m1s6	a12m1s8	a02m4s7	a21m4s3
7	a04f1s4	a23f1s7	a02f1s4	a20f4s5	a07f3s7	a02f3s5	a21f3s5	a01f4s4
8	a09m2s2	a05m3s5	a17m4s1	a13m1s2	a10m2s1	a22m3s8	a18m4s6	a05m4s8
9	a06f3s4	a09f3s7	a18f1s6	a18f2s8	a22f2s4	a08f1s1	a22f4s2	a17f4s8
10	a17m2s5	a11m1s2	a01m4s3	a24m4s6	a20m4s7	a13m2s2	a09m3s4	a04m3s8
11	a23f4s1	a02f4s5	a20f3s2	a17f1s4	a23f3s8	a10f3s2	a11f1s6	a06f1s4
12	a19m4s4	a06m4s8	a15m2s8	a08m4s2	a02m2s5	a07m4s5	a16m2s1	a02m1s4
13	a13f2s6	a24f2s4	a13f4s6	a02f2s7	a16f4s7	a16f1s2	a14f4s7	a07f2s6
14	a08m1s1	a22m4s8	a07m2s7	a14m2s6	a12m4s5	a20m1s4	a10m4s2	a13m4s7
15	a07f4s2	a10f4s6	a10f1s6	a12f4s6	a08f4s2	a11f4s1	a19f1s5	a08f3s7
16	a24m1s2	a03m1s6	a09m4s4	a16m4s7	a18m2s2	a15m4s5	a08m2s3	a10m1s8
17	a21f2s1	a01f3s3	a11f2s4	a01f1s5	a21f1s8	a18f3s1	a03f1s8	a23f2s4
18	a10m3s8	a14m4s7	a22m1s4	a15m3s5	a11m3s4	a04m1s3	a07m1s7	a11m2s5
19	a14f3s6	a15f1s5	a12f3s8	a04f4s1	a24f4s6	a24f1s2	a06f4s6	a16f3s4
20	a03m4s3	a12m2s1	a24m3s4	a23m3s8	a19m3s4	a14m3s4	a24m2s4	a03m2s4
21	a15f4s7	a07f1s7	a04f3s1	a03f3s4	a13f1s2	a17f2s2	a04f2s7	a24f3s4
22	a01m2s6	a13m3s1	a16m3s4	a05m1s5	a09m1s6	a05m2s6	a23m1s5	a18m1s2
23	a22f3s3	a17f3s2	a19f2s7	a09f1s7	a15f3s8	a19f4s1	a20f2s2	a22f1s5
24	a02m3s1	a21m3s4	a14m1s6	a21m1s2	a03m3s4	a21m2s6	a17m3s4	a12m3s3

The randomization lists for each of the eight listening panels for each experiment are provided in /so70/subjctv/exp*/data/play*.lst.

2.3.6 Presentation

Presentation of speech materials for the SO 70 codec listening tests shall be made with one side of high fidelity supra-aural headphones with the other ear uncovered. The speech material delivery system shall meet the requirements of Section 3.3.1.1. The listeners should be seated in a quiet room, with an ambient noise level of 30 dBA or below.

2.3.7 Listeners

The listener sample is intended to represent the population of telephone users with normal hearing acuity. The listeners should be naïve with respect to telephony technology issues; that is, they should not be experts in telephone design, digital voice encoding algorithms, and so on. They should not be trained listeners; that is, they should not have been trained in these or previous listening studies using feedback trials. Age distribution and gender should be nominally balanced across listening panels.

Each listener shall provide data only once for a particular evaluation. A listener may participate in different evaluations, but test sessions performed with the same listener should be at least two months apart so as to reduce the cumulative effects of experience.

2.3.8 Listening Test Procedures

2.3.8.1 ACR Listening Test Procedures – Experiments 1, 3, and 5

The listeners shall listen to each sample and rate the quality of the test sample using a five-point scale, with the points labeled:

5 Excellent

4 Good

3 Fair

2 Poor

1 Bad

Data from 32 listeners shall be used for Experiments 1, 3, and 5, four listeners for each listening panel where each listening panel uses a different randomization. Before starting the test, the listeners should be given instructions for performing the subjective test. An example set of instructions for the ACR are presented in Figure 2.3.8.1-1. The instructions may be modified to allow for variations in laboratory data-gathering apparatus.

This is an experiment to determine the perceived quality of speech over the telephone. You will be listening to a number of recorded speech samples, spoken by several different talkers, and you will be rating how good you think they sound.

Use the single headphone on the ear you normally use for the telephone. On each trial a two-sentence sample will be played. After you have listened to the sample, determine the category from the list below which best describes the overall quality of the sample. Press the numeric key on your keyboard corresponding to your rating for how good or bad that particular passage sounded.

The quality of the speech should be rated according to the scale below:

5 Excellent

4 Good

3 Fair

2 Poor

1 Bad

During the session you will hear samples varying in different aspects of quality. Please take into account your total impression of each sample, rather than concentrating on any particular aspect.

Figure 2.3.8.1-1 Instructions for Listeners

2.3.8.2 P-835 Listening Test Procedures – Experiments 2, 4, and 6

Experiments 2, 4, and 6 use the P.835 test methodology described in ITU-T Rec. P.835 [13]. The P.835 methodology is specifically designed to evaluate the quality of speech in background noise. It yields a measure of Signal Quality (SIG), a measure of Background Quality (BAK), and a measure of Overall Quality (OVRL). In general, OVRL scores are highly correlated with MOS but the OVRL score provides greater sensitivity and precision in test conditions involving background noise. While the OVRL score is of most interest here, the SIG and BAK scores also provide valuable diagnostic information. For each trial in a P.835 test, listeners are presented with three sub-samples where each sub-sample is a single sentence (approx. 4 sec. duration) processed through the same test condition. In one of the first two sub-samples listeners rate the *Signal Quality* on a five-point rating scale with the points labeled:

5 Very natural, no distortion
4 Fairly natural, little distortion
3 Somewhat natural, some distortion
2 Fairly unnatural, fairly distorted
1 Very unnatural, very distorted

For the other of the first two sub-samples listeners rate the *Background Quality* on a five-point rating scale with the points labeled:

5	Not noticeable
4	Fairly noticeable
3	Noticeable but not intrusive
2	Fairly conspicuous, somewhat intrusive
1	Very conspicuous, very intrusive

For the third sub-sample listeners rate the *Overall quality* on a five-point rating scale with the points labeled:

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Data from 32 listeners shall be used for Experiments 2, 4, and 6, four listeners for each listening panel where each listening panel uses a different randomization

Before starting the test, the listeners should be given instructions for performing the subjective test. An example set of instructions for the P.835 test are presented below. The instructions may be modified to allow for variations in laboratory data-gathering apparatus.

Instructions for P.835 Speech Rating Experiment

In this speech rating experiment each trial will involve **three** sentences and you will give a rating for **each** sentence.

For the first sentence in each trial you will be asked to attend **only to the speech signal** and rate how natural, or conversely, how degraded, the **speech signal** sounds to you. You will use the rating scale shown in the figure below to register your ratings of the speech signal. Your task will be to choose the numbered phrase from the list below that best describes your opinion of the **SPEECH SIGNAL ALONE** and then enter the corresponding number on your keyboard.

Attending **ONLY to the SPEECH SIGNAL**, select the category which best describes the sample you just heard.

the **SPEECH SIGNAL** in this sample was

5 – VERY NATURAL, NO DEGRADATION
4 – FAIRLY NATURAL, LITTLE DEGRADATION
3 - SOMEWHAT NATURAL, SOMEWHAT DEGRADED
2 – FAIRLY UNNATURAL, FAIRLY DEGRADED
1 – VERY UNNATURAL, VERY DEGRADED

For the second sentence in each trial you will be asked to attend **only to the background** and rate how noticeable, intrusive, and/or conspicuous the **background** sounds to you. You will use the rating scale shown in the figure below to register your ratings of the background. Your task will be to choose the numbered phrase from the list below that best describes your opinion of the **BACKGROUND ALONE** and then enter the corresponding number on your keyboard.

Attending **ONLY to the BACKGROUND**, select the category which best describes the sample you just heard.

the **BACKGROUND** in this sample was

5 – NOT NOTICEABLE
4 – SOMEWHAT NOTICEABLE
3 – NOTICEABLE BUT NOT INTRUSIVE
2 – FAIRLY CONSPICUOUS, SOMEWHAT INTRUSIVE
1 – VERY CONSPICUOUS, VERY INTRUSIVE

For the third and final sentence in each trial you will be asked to attend to the entire sample (both the speech signal and the background) and rate your opinion of the sample for purposes of everyday speech communication.

Select the category which best describes the sample you just heard for purposes of everyday speech communication.

the **OVERALL SPEECH SAMPLE** was

5 – EXCELLENT
4 – GOOD
3 - FAIR
2 – POOR
1 – BAD

2.3.9 Analysis of Results

The response data from the practice blocks shall be discarded. Data sets with missing responses from listeners shall not be used – i.e., a complete set of data is required for 32 listeners, four for each of eight listening panels. Responses from the different listening panels for the corresponding test conditions shall be treated as equivalent in the analysis.

2.3.9.1 Basic Results for the SO 70 Listening tests

The votes for each of the test conditions for SO 70 Experiments 1 and 3 and Experiments 2 and 4 shall be averaged to produce an associated mean score (***M***) as shown in Equation 2.3.9.1-1 and a Standard Deviation (***SD***) as shown in Equation 2.3.9.1-2, where ***L*** is the number of listeners and ***T*** is the number of talkers involved in the experiment.

$$\overline{M} = \left(\sum_{L} \sum_{T} X_{l,t} \right) \Big/ \left(L \times T \right) \qquad (2.3.9.1\text{-}1)$$

$$SD = \sqrt{\left(\sum_{L} \sum_{T} \left(X_{l,t} - \overline{M} \right)^2 \right) \Big/ \left(L \times T - 1 \right)} \qquad (2.3.9.1\text{-}2)$$

2.3.9.2 Minimum Subjective Requirement for SO 70 Listening Tests

The Terms of Reference for the MPS tests state that the mean score for each of the Test Encoder/Decoder Combinations (E/DC) should be "*not worse than*" the mean score for the Reference E/DC. For most of the test conditions involved in the subjective experiments there are three Test E/DC's (M-T, T-M, and T-T), which means there are three statistical tests against the Reference E/DC (M-M). The three statistical tests are not independent, however. Since they all involve the same ratings for the Reference E/DC, t-tests are not appropriate. The appropriate statistical test for multiple Test conditions against a common Reference condition is Dunnett's Test. A complete description of Dunnett's Test is contained in Appendix B.

The critical value for the Dunnett's test is 2.09 (one-sided test, $p<.05$, 4 E/DC's, df = 93).

For those test conditions where a single Test E/DC (T-T) is compared against the Reference E/DC (M-M), the appropriate statistical test is Student's t-test[5].

The critical value for the Student's t-test is 1.70 (one-sided test, $p<.05$, df = 31).

In both the Dunnett's Test and the t-test the MPS test is evaluated by dividing the difference between the mean score for the Test E/DC and the mean score for the Reference ED/C by the Standard Error of the Mean Difference (SE_{MD}) as shown in Equation 2.3.9.2-1. If the resultant Test value is less than

[5] The appropriate t-test is a "matched groups" t-test and the SE_{MD} is based on the differences between individual listener's average ratings, where the average is over talkers. Therefore, the SE_{MD} is based on 32 difference scores, one for each listener (df =31).

the criterion value for the appropriate test (2.09 for Dunnett's Test, 1.70 for the t-test), then the E/DC passes the MPS test.

$$Test = \left(\overline{M}_{Ref} - \overline{M}_{Test}\right) \Big/ SE_{MD} \qquad (2.3.9.2\text{-}1)$$

2.3.10 Expected Results for Reference Conditions

2.3.10.1 Reference Conditions for Experiments 1, 3, and 5

The MNRU conditions have been included to provide a frame of reference for the Experiments 1, 3, and 5. In listening evaluations where test conditions span approximately the same range of quality, the MOS results for similar conditions should be approximately the same. Data from previous studies allows a generalization to be made concerning the expected MOS results for the MNRU reference conditions (see Figure 2.3.10.1-1).

MOS scores obtained for the MNRU conditions in any SO 70 validation test should be compared to those shown in the graph below. Inconsistencies beyond a small shift in the means in either direction or a slight stretching or compression of the scale near the extremes may imply a problem in the execution of the evaluation test. In particular, MOS should be monotonic with MNRU, within the limits of statistical resolution; and the contour of the relation should show a similar slope.

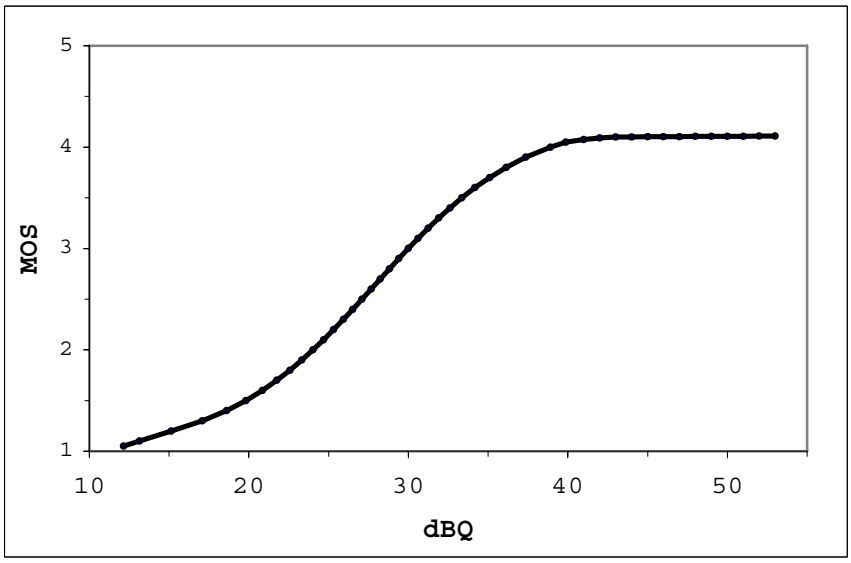


Figure 2.3.10.1-1 Typical Plot of MOS versus MNRU

2.3.10.2 Reference Conditions for Experiments 2, 4, and 6

Reference conditions for P.835 tests are constructed as a combination of SNR and MNRU processing to provide degradation in overall speech quality in two dimensions — signal distortion and background noise intrusiveness. Table 2.3.2.3.2-2 shows the eight reference conditions (b01 – b08)

involved in the P.835 Experiments 2, 4, and 6. In general, results are expected for these reference conditions such that the obtained score profiles are similar to those shown in Figure 2.3.10.2-1.

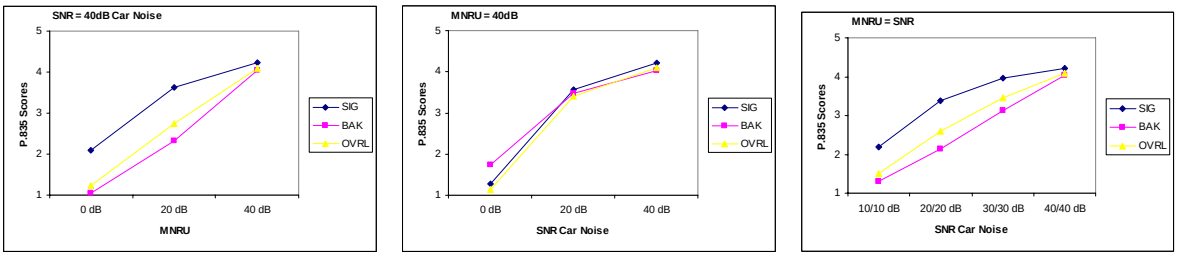


Figure 2.3.10.2-1 Typical P.835 Score Profiles for Reference Conditions

2.4 Performance Testing for SO 73

2.4.1 Objective Performance Testing for SO 73

The objective testing portion of this specification consists of an average data rate test, and compliance to End-to-End Algorithmic Delay and Unity-gain requirements.

2.4.1.1 Average Data Rate Test

An implementation may support SO 73 for 16 kHz sample rates (for example, a mobile station that supports wideband electro-acoustics). The average data rate for the test codec shall be measured using seven source speech files that are contained in the /so73/subjctv/exp{1,2}/source/ directories. Each file exhibits a different condition: power levels: -12 dB, -22 dB, and -32 dB, and background noise conditions: 20 dB SNR babble noise, 10 dB SNR car noise, 20 dB SNR car noise and 15 dB SNR street noise. The input source files used in the average data rate test have an approximate voice activity factor of 0.6, and are the same input files used in the subjective portion of the experiment.

An implementation may support SO 73 only for 8 kHz sample rate input/output (for example, a Base-station transcoder or a Media Gateway or mobile station). The average data rate for the test codec for this case shall be measured using six narrow-band source speech files that are contained in the /so73/subjctv/exp{3,4}/source/ directories. Each file exhibits a different condition: power levels: -12 dB, -22 dB, and -32 dB, and background noise conditions: 20 dB SNR babble noise, 15 dB SNR car noise and 15 dB SNR street noise. The input source files used in the average data rate test have an approximate voice activity factor of 0.6, and are the same input files used in the subjective portion of the experiment.

2.4.1.1.1 Average Data Rate Computation for SO 73

The average channel data rate for the test codec shall be computed for each of the benchmark files as follows:

$$R = (9600*N_1 + 4800*N_2 + 2400*N_4 + 1200*N_8)/N,$$

where

N_1 = number of frames encoded at Rate 1,

N_2 = number of frames encoded at Rate 1/2,

N_4 = number of frames encoded at Rate 1/4,

N_8 = number of frames encoded at Rate 1/8, and

$N = N_1 + N_2 + N_4 + N_8$.

For the 16 kHz input:

The total average channel data rate for the test codec is then given by:

R_{avg} = 1/7 * { R(ambient background segment @ -12dB).+ R(ambient background segment @ -32dB).+ R(ambient background segment @ -22dB). + R(20 dB SNR babble noise segment @ -22dB) + R(10 dB SNR car noise segment @ -22dB). + R(20 dB SNR car noise segment @ -22dB).+ R(15 dB SNR street noise segment @ -22dB) }.

For the 8 kHz input:

The total average channel data rate for the test codec is then given by:

R_{avg} = 1/6 * { R(ambient background segment @ -12dB).+ R(ambient background segment @ -32dB).+ R(ambient background segment @ -22dB). + R(20 dB SNR babble noise segment @ -22dB) + R(15 dB SNR car noise segment @ -22dB). + R(15 dB SNR street noise segment @ -22dB) }.

The above files are to be processed with EVRC-NW encoder at various capacity operating points (defined by the active speech average channel rate) shown in Table 2.4.1.1.1-1.

Table 2.4.1.1.1-1 Target ADR vs Capacity Operating Point

Capacity Operating Point (active speech average channel data rate)	Target Average Channel Data Rate, kbps
EVRC-NW - RATE_REDUC='000'	5.6+1.5%
EVRC-NW - RATE_REDUC='001'	5.92+1.5%
EVRC-NW - RATE_REDUC='010'	4.82+1.5%
EVRC-NW - RATE_REDUC='011'	4.57+1.5%
EVRC-NW - RATE_REDUC='100'	4.259+1.5%
EVRC-NW - RATE_REDUC='101'	4.079+1.5%
EVRC-NW - RATE_REDUC='110'	3.954+1.5%
EVRC-NW - RATE_REDUC='111'	3.29+1.5%

The above table provides the maximum allowable average channel rate (including full, half, quarter and eighth-rate) for the different operating points. These maximum allowable average channel rates were obtained by processing the 7 wide-band benchmark files for the 16 kHz case and 6 narrow-band benchmark files for the 8kHz case through the master floating point software. See Section 3.4.2.1 for details in using the provided software tool that can be used to aid in making this calculation.

2.4.1.1.2 Average Data Rate Requirement for SO 73

The total average data rate R_{avg} for each operating point shall not exceed the target average data rate by more than the tolerance level in Table 2.4.1.1.1-1, otherwise the test codec fails the compliance test.

2.4.1.2 Unity Gain Requirement

The specific EVRC-NW test codec shall output speech with unity gain when compared with the input speech. The unity gain measurement (output active–speech level/input active speech level) will be performed over the entire input speech database for the clean, nominal-level source conditions for each mode. The measurement should be made using the STL-2000 tool [6] [6a] actlev, and must not show more than $\pm$ 0.5 dB deviation between input and output active speech levels. This procedure is fully described in [9].

2.4.1.3 End-to-end Algorithmic Delay Recommendation

The algorithmic delay for the specific EVRC-NW test codec should be calculated analytically by the codec manufacturer. In considering the algorithmic delay, it can be assumed that all transmission channels have infinite bandwidth, and that all processing elements have infinite throughput. Algorithmic delay is defined as the sum of all sequential filter delays and buffering delays in the encode/decode path.

The maximum end-to-end algorithmic delay should be no greater than that of the master codec. For the master codecs defined in [1], the algorithmic delay is given as:

Delay Source	Delay (ms)
Signal Preprocessing Delay:	0.0
Filterbank Analysis	0.8
LPC Analysis “Look-ahead”:	10.0
LPC Analysis Window:	20.0
Highband excitation generation delay	1.5
Highband synthesis overlap-and-add delay	2.0
Filterbank Synthesis Delay:	1.1
Total:	35.4

Therefore, the total algorithmic delay imposed by a SO 73 test codec should not exceed 35.4 milliseconds.

2.4.2 Subjective Performance Testing for SO 73

This section outlines the subjective testing methodology of the subjective performance test. The purpose of this testing is to evaluate the quality of the test codec under a variety of conditions which may occur in the CDMA system. To accomplish this, suites of listening experiments have been designed to test speech codec quality under a variety of conditions depending on a number of parameters. These conditions include channel impairments, audio background noise, and different input levels.

Figure 2.4.2-1 illustrates a decision tree to arrive at the suite of tests that are needed to demonstrate Minimum Performance Spec compliance of a Test implementation of SO 73 for different profiles of equipment that support SO 73.

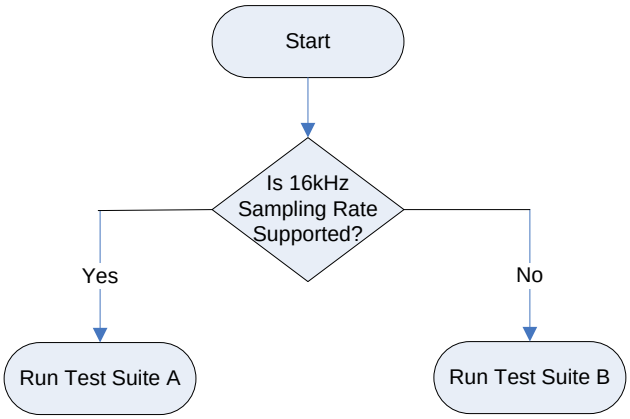


Figure 2.4.2-1 SO 73 Subjective test suite decision flowchart

An implementation may support SO 73 only for 8 kHz sample rate input/output (for example, a Base-station transcoder or a Media Gateway). An implementation may support SO 73 for both 16 kHz and 8 kHz sample rates (for example, a mobile station that supports wideband electro-acoustics).

Therefore, the main parameter in the decision tree is 16 kHz support in the implementation.

Depending on the implementation profile of the device under test, one of 2 possible Test Suites are to be used to demonstrate SO 73 compliance. These 2 test suites named Test suites A and B, and the individual tests comprising the Test suites are highlighted in Table 2.4.2-1

Table 2.4.2-1 Test Suites for SO 73 compliance

Test Suites	Set of Experiments
A	Experiment 1, 2, 3, 4, 5 and 6
B	Experiment 3, 4, and 6

Each of the individual experiments are further defined in detail by Table 2.4.2-2.

Table 2.4.2-2 Experiments for SO 73 compliance

Experiment	Individual tests	Notes
1	WB clean/level/FER/signaling - ACR	Mobile supporting 16 kHz Fs
2	WB noise/FER - P.835	Mobile supporting 16 kHz Fs
3	NB clean/level/FER/signaling including SO 68 interoperable mode tests - ACR	BS supporting 8 kHz, and MS supporting 8/16 kHz)
4	NB noise/FER including SO 68 interoperable mode tests - P.835	BS supporting 8 kHz, and MS supporting 8/16 kHz)
5	WB music decoder test – ACR	Mobile supporting 16 kHz Fs
6	NB music decoder test – ACR	BS supporting 8 kHz Fs

2.4.2.1 Definition

The codec subjective test is intended to validate the implementation of the speech codec being tested using the master codec defined in 3.4.3 as a reference. Experiments 1 and 3 are based on the Absolute Category Rating (ACR) method, which yields the Mean Opinion Score (MOS) as described in [10]. Experiments 2 and 4 are based on the ITU-T Recommendation P.835 described in [13].

2.4.2.2 Method of Measurement

The subjective tests involve a listening-only assessment of the quality of the codec being tested, using the master codec as a reference. Subjects from the general population of telephone users will rate the various conditions of the test. Material supplied with this standard for use with this test includes source speech, impaired packet files from the master codec encoder, and source speech processed by various Modulated Noise Reference Unit (MNRU) conditions and other references. The basic Absolute Category Rating test procedure involves rating all conditions using a five-point scale describing the opinion of the test condition. This procedure is fully described in [10]. The P.835 test method involves rating all conditions on scales of “Signal”, “Background”, and “Overall” quality and is fully described in [13].

2.4.2.3 Test Conditions and Test Design for SO 73

Listening experiments 1 and 3 for SO 73 are performed as ACR listening tests. Experiments 2 and 4 for SO 73 are performed as P.835 listening tests.

2.4.2.3.1 Subjective Experiment 1 for SO 73

The Test Parameters for Listening Experiment 1 are presented in Table 2.4.2.3.1-1.

Table 2.4.2.3.1-1 SO 73 Listening Experiment 1 Test Parameters

Condition	Description
Type of test	MOS (P.800), Wideband
Number of talkers	4 males, 4 females
Background noise	none (ambient)
Audio Input Level	-22 dB, -32 dB, -12 dB
Filter characteristics	P.341 (refer Section 3.3.2.4)
Reference conditions	(8) Specified reference conditions
Test conditions	o Low Audio Input Level -32 dB + 1% d&b o Nominal Audio Input Level, -22 dB o High Audio Input Level -12 dB o 3% FER and 1%FER + 2%pls at Nominal Audio Input Level, -22
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

The Test Conditions for Listening Experiment 1 are presented in Table 2.3.2.3.1-2.

Table 2.4.2.3.1-2 SO 73 Listening Experiment 1 Test Conditions

Exp.1	Wideband - ACR	
Reference Conditions		
File	**MNRU**	
a01	7dB MNRU	Reference
a02	14dB MNRU	Reference
a03	21dB MNRU	Reference
a04	28dB MNRU	Reference
a05	35dB MNRU	Reference
a06	42dB MNRU	Reference
a07	49dB MNRU	Reference
a08	Direct Source	Reference
Test Conditions		
File	**Condition**	**Enc-Dec**
a09	Nominal level	M-M
a10	Nominal level	M-T
a11	Nominal level	T-T
a12	Nominal level	T-M
a13	Low level, 1% d&b	M-M
a14	Low level, 1% d&b	M-T
a15	Low level, 1% d&b	T-T
a16	Low level, 1% d&b	T-M
a17	High level	M-M
a18	High level	M-T
a19	High level	T-T
a20	High level	T-M
a21	1% FER, 1% PLS	M-M
a22	1% FER, 1% PLS	M-T
a23	3% FER	M-M
a24	3% FER	M-T

2.4.2.3.2 Subjective Experiment 2 for SO 73

The Test Parameters for Listening Experiment 2 are presented in Table 2.3.2.3.2-1.

Table 2.4.2.3.2-1 SO 74 Listening Experiment 2 Test Parameters

Condition	Description
Type of test	P-NSA (P.835), Wideband
Number of talkers	3 males, 3 females
Background noise	Specified test conditions
Audio Input Level	-22 dB
Filter characteristics	P.341 (refer Section 3.3.2.4)
Reference conditions	(8) Specified reference conditions
Test conditions	o Car Noise, 10 dB SNR o Car Noise, 20 dB SNR + 2% FER o Street Noise, 15 dB SNR o Babble noise, 20 dB S/N
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

The Test Conditions for Listening Experiment 2 are presented in Table 2.4.2.3.2-2.

Table 2.4.2.3.2-2 SO 73 Listening Experiment 2 Test Conditions.

Exp.2	**Wideband - P.835**	
Reference Conditions		
File	**MNRU, SNR**	
b01	MNRU=40dB, SNR=40dB	Reference
b02	MNRU=40dB, SNR=20dB	Reference
b03	MNRU=40dB, SNR=0dB	Reference
b04	MNRU=0dB, SNR=40dB	Reference
b05	MNRU=20dB, SNR=40dB	Reference
b06	MNRU=10dB, SNR=10dB	Reference
b07	MNRU=20dB, SNR=20dB	Reference
b08	MNRU=40dB, SNR=30dB	Reference
Test Conditions		
File	**Condition**	**Enc-Dec**
b09	Car 10dB SNR	M-M
b10	Car 10dB SNR	M-T
b11	Car 10dB SNR	T-T
b12	Car 10dB SNR	T-M

Exp.2	Wideband - P.835	
b13	Car 20dB SNR + 2% FER	M-M
b14	Car 20dB SNR + 2% FER	M-T
b15	Car 20dB SNR + 2% FER	T-T
b16	Car 20dB SNR + 2% FER	T-M
b17	Street 15dB SNR	M-M
b18	Street 15dB SNR	M-T
b19	Street 15dB SNR	T-T
b20	Street 15dB SNR	T-M
b21	Babble 20dB SNR	M-M
b22	Babble 20dB SNR	M-T
b23	Babble 20dB SNR	T-T
b24	Babble 20dB SNR	T-M

2.4.2.3.3 Subjective Experiment 3 for SO 73

The Test Parameters for Listening Experiment 3 are presented in Table 2.4.2.3.3-1.

Table 2.4.2.3.3-1 SO 73 Listening Experiment 3 Test Parameters

Condition	Description
Type of test	ACR (P.800), Narrowband
Number of talkers	4 males, 4 females
Background noise	none (ambient)
Audio Input Level	-22 dB, -32 dB, -12 dB
Filter characteristics	MIRS
Reference conditions	(8) Specified reference conditions
Test conditions	o Nominal level, Modes 0, 4, 7 o Low level, Modes 0, 4 o High Level, Mode 0, 4 o 1% d&b, 1% pls, Modes 0, 4 o 3% FER, Modes 0, 4
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

The Test Conditions for Listening Experiment 3 are presented in Table 2.4.2.3.3-2**.**

Table 2.4.2.3.3-2 SO 73 Listening Experiment 3 Test Conditions.

Label	Operating Point	Condition	Encoder/Decoder Combinations
c01	Reference	MNRU 5dB	
c02	Reference	MNRU 10dB	
c03	Reference	MNRU 15dB	
c04	Reference	MNRU 20dB	
c05	Reference	MNRU 25dB	
c06	Reference	MNRU 30dB	
c07	Reference	MNRU 35dB	
c08	Reference	Direct	
c09	Mode 1	Nominal, -22 dB	M-M
c10	Mode 1	Nominal, -22 dB	M-T
c11	Mode 1	Nominal, -22 dB	T-T
c12	Mode 1	Nominal, -22 dB	T-M
c13	Mode 6	Nominal, -22 dB	M-M
c14	Mode 6	Nominal, -22 dB	M-T
c15	Mode 6	Nominal, -22 dB	T-T
c16	Mode 6	Nominal, -22 dB	T-M
c17	Mode 7	Nominal, -22 dB	M-M
c18	Mode 7	Nominal, -22 dB	M-T
c19	Mode 7	Nominal, -22 dB	T-T
c20	Mode 7	Nominal, -22 dB	T-M
c21	Mode 1	Low, -32 dB, 1% d&b, 1% pls	M-M
c22	Mode 1	Low, -32 dB, 1% d&b, 1% pls	M-T
c23	Mode 1	Low, -32 dB, 1% d&b, 1% pls	T-T
c24	Mode 1	Low, -32 dB, 1% d&b, 1% pls	T-M
c25	Mode 6	Low, -32 dB, 1% d&b, 1% pls	M-M
c26	Mode 6	Low, -32 dB, 1% d&b, 1% pls	M-T
c27	Mode 6	Low, -32 dB, 1% d&b, 1% pls	T-T
c28	Mode 6	Low, -32 dB, 1% d&b, 1% pls	T-M
c29	Mode 1	High, -12 dB	M-M
c30	Mode 1	High, -12 dB	M-T
c31	Mode 1	High, -12 dB	T-T
c32	Mode 1	High, -12 dB	T-M
c33	Mode 6	High, -12 dB	M-M
c34	Mode 6	High, -12 dB	M-T
c35	Mode 6	High, -12 dB	T-T
c36	Mode 6	High, -12 dB	T-M

Label	Operating Point	Condition	Encoder/Decoder Combinations
c37	Mode 1	Nominal, -22 dB, 3% FER	M-M
c38	Mode 1	Nominal, -22 dB, 3% FER	M-T
c39	Mode 6	Nominal, -22 dB, 3% FER	M-M
c40	Mode 6	Nominal, -22 dB, 3% FER	M-T
c41	Mode 0	Nominal, LB portion of Widemode mode – decoder test only	M-M
c42	Mode 0	Nominal, LB portion of Widemode mode – decoder test only	M-T
c43	Mode 0	Low, 1% D&B, 1% PLS, LB portion of Widemode mode – decoder test only	M-M
c44	Mode 0	Low, 1% D&B, 1% PLS, LB portion of Widemode mode – decoder test only	M-T
c45	Mode 0	High, LB portion of Widemode mode – decoder test only	M-M
c46	Mode 0	High, LB portion of Widemode mode – decoder test only	M-T
c47	Mode 0	Nominal, 3% FER, LB portion of Widemode mode – decoder test only	M-M
c48	Mode 0	Nominal, 3% FER, LB portion of Widemode mode – decoder test only	M-T

2.4.2.3.4 Subjective Experiment 4 for SO 73

The Test Parameters for Listening Experiment 4 are presented in Table 2.4.2.3.4-1.

Table 2.4.2.3.4-1 SO 74 Listening Experiment 4 Test Parameters

Condition	Description
Type of test	P-NSA (P.835), Narrowband
Number of talkers	3 males, 3 females
Background noise	Specified test conditions
Audio Input Level	-22 dB
Filter characteristics	MIRS
Reference conditions	(8) Specified reference conditions
Test conditions	o Car Noise, 15 dB SNR, Modes 0, 4, 7 o Street Noise, 15 dB SNR, Modes 0, 4 o Babble noise, 20 dB SNR, 2%FER, Modes 0, 4
Encoder/Decoder Combinations	(4) M/M, M/T, T/T, T/M

The Test Conditions for Listening Experiment 4 are presented in Table 2.4.2.3.4-2.

Table 2.4.2.3.4-2 SO 73 Listening Experiment 4 Test Conditions.

Exp.4	Narrowband - P.835		
Reference Conditions			
File	**MNRU**		
d01	MNRU=40dB, SNR=40dB		Reference
d02	MNRU=40dB, SNR=20dB		Reference
d03	MNRU=40dB, SNR=0dB		Reference
d04	MNRU=0dB, SNR=40dB		Reference
d05	MNRU=20dB, SNR=40dB		Reference
d06	MNRU=10dB, SNR=10dB		Reference
d07	MNRU=20dB, SNR=20dB		Reference
d08	MNRU=40dB, SNR=30dB		Reference
Test Conditions			
File	**Condition**		**Enc-Dec**
d09	Car 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-M
d10	Car 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-T
d11	Car 15dB SNR, Mode 1	(interoperable with Mode 0 of SO 68) support	M-M
d12	Car 15dB SNR, Mode 1	(interoperable with Mode 0 of SO 68) support	M-T
d13	Car 15dB SNR, Mode 1	(interoperable with Mode 0 of SO 68) support	T-T
d14	Car 15dB SNR, Mode 1	(interoperable with Mode 0 of SO 68) support	T-M
d15	Car 15dB SNR, Mode 7	(interoperable with Mode 0 of SO 68) support	T-T
d16	Car 15dB SNR, Mode 7	(interoperable with Mode 0 of SO 68) support	T-M
d17	Car 15dB SNR, Mode 7	(interoperable with Mode 0 of SO 68) support	M-M
d18	Car 15dB SNR, Mode 7	(interoperable with Mode 0 of SO 68) support	M-T
d19	Street 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-M
d20	Street 15dB SNR, Mode 0	LB portion of Wideband mode - decoder test only	M-T
d21	Street 15dB SNR, Mode 1	(interoperable with Mode 0 of SO 68) support	M-M
d22	Street 15dB SNR, Mode 1	(interoperable with Mode 0 of SO 68) support	M-T
d23	Street 15dB SNR, Mode 1	(interoperable with Mode 0 of SO 68) support	T-T
d24	Street 15dB SNR, Mode 1	(interoperable with Mode 0 of SO 68) support	T-M

Exp.4	Narrowband - P.835		
d25	Babble 20dB SNR, 2% FER, Mode 0	LB portion of Wideband mode - decoder test only	M-M
d26	Babble 20dB SNR, 2% FER, Mode 0	LB portion of Wideband mode - decoder test only	M-T
d27	Babble 20dB SNR, 2% FER, Mode 1	(interoperable with Mode 0 of SO 68) support	M-M
d28	Babble 20dB SNR, 2% FER, Mode 1	(interoperable with Mode 0 of SO 68) support	M-T
d29	Babble 20dB SNR, 2% FER, Mode 1	(interoperable with Mode 0 of SO 68) support	T-T
d30	Babble 20dB SNR, 2% FER, Mode 1	(interoperable with Mode 0 of SO 68) support	T-M
d31	Car 15dB SNR, Mode 6	(interoperable with Mode 6 of SO 68) support	M-M
d32	Car 15dB SNR, Mode 6	(interoperable with Mode 6 of SO 68) support	M-T
d33	Car 15dB SNR, Mode 6	(interoperable with Mode 6 of SO 68) support	T-T
d34	Car 15dB SNR, Mode 6	(interoperable with Mode 6 of SO 68) support	T-M
d35	Street 15dB SNR, Mode 6	(interoperable with Mode 6 of SO 68) support	M-M
d36	Street 15dB SNR, Mode 6	(interoperable with Mode 6 of SO 68) support	M-T
d37	Street 15dB SNR, Mode 6	(interoperable with Mode 6 of SO 68) support	T-T
d38	Street 15dB SNR, Mode 6	(interoperable with Mode 6 of SO 68) support	T-M
d39	Babble 20dB SNR, 2% FER, Mode 6	(interoperable with Mode 6 of SO 68) support	M-M
d40	Babble 20dB SNR, 2% FER, Mode 6	(interoperable with Mode 6 of SO 68) support	M-T
d41	Babble 20dB SNR, 2% FER, Mode 6	(interoperable with Mode 6 of SO 68) support	T-T
d42	Babble 20dB SNR, 2% FER, Mode 6	(interoperable with Mode 6 of SO 68) support	T-M

2.4.2.3.5 Subjective Experiment 5 for SO 73

The Test Parameters for Listening Experiment 5 are presented in Table 2.4.2.3.5-1.

Table 2.4.2.3.5-1 SO 73 Listening Experiment 5 Test Parameters

Condition	Description
Type of test	ACR (P.800), Wideband
Number of genres	4
Background noise	none (ambient)
Audio Input Level	-22 dB
Filter characteristics	P.341 (refer Section 3.3.2.4)
Reference conditions	(4) Specified reference conditions
Test conditions	0% FER and 3% FER
Encoder/Decoder Combinations	(2) M/M, M/T

The Test Conditions for Listening Experiment 5 are presented in Table 2.4.2.3.5-2.

Table 2.4.2.3.5-2 SO 73 Listening Experiment 5 Test Conditions.

Exp. 5	Wideband Music	
File	**Reference Condition**	
e01	MNRU=15dB	Reference
e02	MNRU=25dB	Reference
e03	MNRU=35dB	Reference
e04	Source	Reference
File	**Test Condition**	**Enc-Dec**
e05	0% FER	M-M
e06	0% FER	M-T
e07	3% FER	M-M
e08	3% FER	M-T

2.4.2.3.6 Subjective Experiment 6 for SO 73

The Test Parameters for Listening Experiment 6 are presented in Table 2.4.2.3.6-1.

Table 2.4.2.3.6-1 SO 73 Listening Experiment 6 Test Parameters

Condition	Description
Type of test	ACR (P.800), Narrowband
Number of genres	4
Background noise	none (ambient)
Audio Input Level	-22 dB
Filter characteristics	MIRS
Reference conditions	(4) Specified reference conditions
Test conditions	0% FER and 3% FER
Encoder/Decoder Combinations	(2) M/M, M/T

The Test Conditions for Listening Experiment 6 are presented in Table 2.4.2.3.6-2.

Table 2.4.2.3.6-2 SO 73 Listening Experiment 6 Test Conditions.

Exp. 6	Narrowband Music	
File	**Reference Condition**	
f01	MNRU=10dB	Reference
f02	MNRU=20dB	Reference
f03	MNRU=30dB	Reference
f04	Source	Reference
File	**Test Condition**	**Enc-Dec**
f05	0% FER	M-M
f06	0% FER	M-T
f07	3% FER	M-M
f08	3% FER	M-T

2.4.2.3.7 Numerical Parameters for the SO 73 Listening Experiments

Table 2.4.2.3.7-1 describes the resultant numerology that is used for the six SO 73 listening experiments. The first column is the description of the parameter and columns 2 to 7 show the numerical value for each of the parameters, for the six listening experiments. For each listening experiment, the different Encode/Decode Test conditions include various interconnections between the Master and Test Encoders, and the Master and Test Decoders. There are eight reference conditions in each of the experiments 1 through 4, and four reference conditions in experiments 5 and 6.

Table 2.4.2.3.7-1 Numerical Parameters for the SO 73 Listening Experiments

Parameter	Exp.1	Exp.2	Exp.3	Exp.4	Exp.5	Exp.6
Type of test	ACR	P.835	ACR	P.835	ACR	ACR
Encode/Decode Test conditions	16	16	40	34	4	4
Reference Conditions	8	8	8	8	4	4
Total Conditions	24	24	48	42	8	8
Talkers (* genres)	8	6	8	6	3*	3*
Stimuli per Talker (* genres)	8	8	8	8	4*	4*
Stimuli per Condition	64	48	64	48	12	12
Total Stimuli per Experiment	1536	1152	3072	2016	96	96
Listening Panels	8	8	8	8	4	4
Stimuli per Listening Panel	192	144	384	252	24	24

Parameter	Exp.1	Exp.2	Exp.3	Exp.4	Exp.5	Exp.6
Listeners (Voters)	32	32	32	32	32	32
Listeners (Voters) per Listening Panel	4	4	4	4	8	8
Votes per Condition	256	192	256	192	96	96

2.4.3 Speech Material for SO 73 Testing

The source speech files used for SO 73 compliance testing consist of Harvard sentences pairs, which are preprocessed to include proper level adjustment and noise mixing for use in the subjective experiments. The talkers used in these files consist of adult males and adult females, and are native speakers of North American English.

For the following discussion, it may be useful to refer to Table 4-3 for the composition of the Software Distribution database.

The source speech material for subjective Experiments is contained in directory /so73/subjctv/exp*/source. Each file has been appropriately pre-filtered, level adjusted, and noise-processed. These files are named src.*. The speech database also includes samples processed through the various reference conditions in directory /so73/subjctv/exp*/ref. The reference conditions are named ref.* for the respective conditions given in the tables in Section 2.4.2.3.

2.4.4 Processing of Speech Material for SO 73 Testing

The source speech material shall be processed by the various combinations of encoders and decoders listed in the descriptions of the experiments given in Section 2.4.2. The master codec software described in Section 3.4.3 shall be used in the processing involving the master codec. Generally, the master codec encoder and decoder outputs have been provided in the respective directories, /so73/subjctv/exp*/m_pkt and /so73/subjctv/exp*/m_m. Execution of the master codec software is needed only for the test encoder/master decoder combination for each experiment/condition.

All codec processing shall be done digitally. Noise suppression and post-filter options shall be enabled for both the master and the test codecs. The digital format of the speech files is described in Section 3.4.4.4.

The naming convention of the processed speech is as follows: For the packet files in the /so73/subjctv/exp{1,3}/m_pkt directory, the *.p12 files are the master packet files for the *.s12 source file. Likewise, the *.p22 and *.p32 files are the respective packet files for the *.s22 and *.s32 source files. The *.pf3 files are the impaired packet files which will be described in Section 2.4.4.3.

Similarly, the directory /so73/subjctv/exp{2,4}/m_pkt contains the master packet files for the respective experiments. Here, the *.pc10, *.pb20, and *.ps files are the master packet files for the *.c15, *.b20, and *.s15 source files, respectively.

For the master encode/master decode directories (/so73/subjctv/exp*/m_m), the naming convention of the speech files is such that the first two characters of the file name indicate the codec combination and the suffix indicates the condition numbers in Table 2.4.2.3.1-2 and Table 2.4.2.3.2-2.

Naming conventions for the remaining two experiments follow accordingly.

Detailed descriptions of all processing operations are given in Section 6.

2.4.4.1 Encoding by the Test Codec

All of the source files will be encoded by the test codec to produce encoded packet files. For ease of reference, it is recommended that directories /so73/subjctv/exp*/t_pkt be created to deposit the test encoder output packets, and that the naming conventions be made consistent with the master codec.

2.4.4.2 Decoding by the Master/Test Codecs

The encoded packet files generated from the various encoders/conditions shall be processed through the master and test decoders.

2.4.4.3 Introduction of Impairments

For the frame error conditions, the impaired master codec encoded packet files are provided in the /so73/subjctv/exp*/m_pkt directory. Unlike other conditions, this condition uses only the test decoder and not the test encoder.

For the Dim-and-Burst processing, and also the Packet Level Signaling conditions, the processing requires inputs from a signaling file to control maximum encoding rate. An external software utility (Evrc_nw_iwf in Section 3.4.2.3) is also needed to reduce the data rate of certain packets from full rate to half rate. Details of these operations are given in Section 6. The signaling file and other utilities are provided in /so73/tools/ directory.

2.4.4.4 Ensuring Proper Encoded Frame Packet Files

All encoded frame packet files shall be examined to ensure that the files only contain data in those file locations where data should exist for a given data rate.

The examination of the encoded frame packet files should indicate the occurrence of any improper data in the files but the examination must not alter the encoded frame packet files in any way.

2.4.4.5 Post-processing of test-condition output files

In order to build the play sets to be presented to the listening panels, the output files for the various test conditions must be processed to provide the appropriate listening conditions. In addition, the concatenated output files must be partitioned into the samples representing the combination of test-condition and talker. The listening conditions for Narrowband experiments are provided by filtering the output files using the STL software tool (*filter*) with the MIRS-receive filter mask. The listening conditions for Wideband experiments are provided by mixing (STL tool *oper*) the output files with Psophometrically filtered noise (STL tool *filter*, PSO filter mask) at 74dBov. STL tool *astrip* is also used to split the concatenated files into the individual samples appropriate for the experiment. Table 2.4.4.5-1 shows the cutting-points to be used with the *astrip* tool for producing the two-sentence samples for the Experiments 1 and 3 ACR test. Table 2.4.4.5-2 shows the cutting-points to be used with the *astrip* tool for producing the single-sentence sub-samples for the Experiments 2 and 4 P.835 test. Table 2.4.4.5-3 shows the sub-samples that make up the samples (i.e., sentence triads) for the P.835 test.

Table 2.4.4.5-1 Cutting Points for the astrip Software Tool for the SO 73 Experiments 1 and 3 ACR Test

Experiment I - ACR							
Sentence-pair		Start sample	Length (samples)	Sentence-pair		Start sample	Length (samples)
1	m1p1	1	113706	33	m1p5	3615102	113446
2	f1p1	113707	118586	34	f1p5	3728548	114249
3	m2p1	232293	111900	35	m2p5	3842797	115062
4	f2p1	344193	117486	36	f2p5	3957859	117344
5	m3p1	461679	110993	37	m3p5	4075203	129258
6	f3p1	572672	123570	38	f3p5	4204461	117851
7	m4p1	696242	106749	39	m4p5	4322312	105606
8	f4p1	802991	110876	40	f4p5	4427918	111339
9	m1p2	913867	102934	41	m1p6	4539257	111723
10	f1p2	1016801	120088	42	f1p6	4650980	105818
11	m2p2	1136889	124661	43	m2p6	4756798	109458
12	f2p2	1261550	111121	44	f2p6	4866256	122664
13	m3p2	1372671	110603	45	m3p6	4988920	116439
14	f3p2	1483274	126079	46	f3p6	5105359	127468
15	m4p2	1609353	99074	47	m4p6	5232827	109566
16	f4p2	1708427	108801	48	f4p6	5342393	108807
17	m1p3	1817228	102960	49	m1p7	5451200	118850
18	f1p3	1920188	118392	50	f1p7	5570050	111097
19	m2p3	2038580	121905	51	m2p7	5681147	121218
20	f2p3	2160485	120916	52	f2p7	5802365	116957
21	m3p3	2281401	104536	53	m3p7	5919322	112149
22	f3p3	2385937	108073	54	f3p7	6031471	110715
23	m4p3	2494010	97510	55	m4p7	6142186	101539
24	f4p3	2591520	107375	56	f4p7	6243725	118876
25	m1p4	2698895	116301	57	m1p8	6362601	118866
26	f1p4	2815196	105011	58	f1p8	6481467	116617
27	m2p4	2920207	124711	59	m2p8	6598084	130938
28	f2p4	3044918	115760	60	f2p8	6729022	123975
29	m3p4	3160678	119447	61	m3p8	6852997	115326
30	f3p4	3280125	110386	62	f3p8	6968323	121531
31	m4p4	3390511	111967	63	m4p8	7089854	104458
32	f4p4	3502478	112624	64	f4p8	7194312	102903

Table 2.4.4.5-2 Cutting Points for the *astrip* Software Tool for the SO 73 Experiments 2 and 4 P.835 Test

Experiment II - P.835											
Sentence		Start sample	Length (samples)	Sentence		Start sample	Length (samples)	Sentence		Start sample	Length (samples)
1	m1s01	1	57758	33	m3s05	1853425	56706	65	m2s11	3679398	55140
2	m1s02	57759	54087	34	m3s06	1910131	49310	66	m2s12	3734538	55821
3	f1s01	111846	58737	35	f3s05	1959441	55546	67	f2s11	3790359	61656
4	f1s02	170583	60083	36	f3s06	2014987	50615	68	f2s12	3852015	60100
5	m2s01	230666	57758	37	m1s07	2065602	62653	69	m3s11	3912115	55432
6	m2s02	288424	56412	38	m1s08	2128255	54048	70	m3s12	3967547	59225
7	f2s01	344836	59226	39	f1s07	2182303	50470	71	f3s11	4026772	64087
8	f2s02	404062	59226	40	f1s08	2232773	54338	72	f3s12	4090859	63406
9	m3s01	463288	51884	41	m2s07	2287111	60526	73	m1s13	4154265	62290
10	m3s02	515172	59593	42	m2s08	2347637	62846	74	m1s14	4216555	55664
11	f3s01	574765	65834	43	f2s07	2410483	56078	75	f1s13	4272219	56276
12	f3s02	640599	54943	44	f2s08	2466561	62846	76	f1s14	4328495	55460
13	m1s03	695542	47316	45	m3s07	2529407	60236	77	m2s13	4383955	64737
14	m1s04	742858	57323	46	m3s08	2589643	55208	78	m2s14	4448692	57193
15	f1s03	800181	58954	47	f3s07	2644851	56465	79	f2s13	4505885	61169
16	f1s04	859135	59389	48	f3s08	2701316	55691	80	f2s14	4567054	55052
17	m2s03	918524	64366	49	m1s09	2757007	57238	81	m3s13	4622106	54440
18	m2s04	982890	58954	50	m1s10	2814245	55305	82	m3s14	4676546	55970
19	f2s03	1041844	54821	51	f1s09	2869550	62506	83	f3s13	4732516	58213
20	f2s04	1096665	60042	52	f1s10	2932056	53308	84	f3s14	4790729	53013
21	m3s03	1156707	51558	53	m2s09	2985364	61148	85	m1s15	4843742	56582
22	m3s04	1208265	57758	54	m2s10	3046512	54353	86	m1s16	4900324	61688
23	f3s03	1266023	64393	55	f2s09	3100865	59266	87	f1s15	4962012	59292
24	f3s04	1330416	57758	56	f2s10	3160131	61148	88	f1s16	5021304	59083
25	m1s05	1388174	59172	57	m3s09	3221279	59580	89	m2s15	5080387	65752
26	m1s06	1447346	47569	58	m3s10	3280859	64388	90	m2s16	5146139	63251
27	f1s05	1494915	55981	59	f3s09	3345247	59371	91	f2s15	5209390	64398
28	f1s06	1550896	60912	60	f3s10	3404618	58012	92	f2s16	5273788	60646
29	m2s05	1611808	61492	61	m1s11	3462630	54265	93	m3s15	5334434	54602
30	m2s06	1673300	61492	62	m1s12	3516895	56113	94	m3s16	5389036	60125
31	f2s05	1734792	57721	63	f1s11	3573008	54459	95	f3s15	5449161	59083
32	f2s06	1792513	60912	64	f1s12	3627467	51931	96	f3s16	5508244	61792

Table 2.4.4.5-3 Composition of the Sentence-Triad Samples for the Experiments 2 and 4 P.835 Test

Sentence-triad	Sentence 1	Sentence 2	Sentence 3
t1	s01	s02	s03
t2	s04	s05	s06
t3	s07	s08	s09
t4	s10	s11	s12
t5	s13	s14	s15
t6	s16	s01	s02
t7	s03	s04	s05
t8	s06	s07	s08

2.4.5 Randomization

For each of the first four subjective experiments, each presentation sample consists of a speech sample processed under a condition of the test. For the ACR Experiments 1 and 3 the sample consists of a pair of concatenated sentences of approximately 8 sec. duration. For the P.835 Experiments 2 and 4 the sample consists of three sub-samples, where each sub-sample is a single sentence of approximately 4 sec. duration. The samples shall be presented to the listeners in a

randomized presentation order. The listeners for each file set shall be presented with practice trials for subjective Experiments 1 and 3 and Experiments 2 and 4. The randomization of the test samples has been accomplished with the following constraints for each of the two experiments:

1. A trial, i.e., a test sample, for the combination of each test condition and each talker shall be presented exactly once to each listening panel (i.e., # trials/panel = # conditions x # talkers).
2. Randomization is in “blocks”, such that one sample of each test condition is presented once, with a randomly selected talker, in each block. This ensures that listeners rate each test condition equally often in the initial, middle and final parts of the block and controls for the effects of time and order of presentation. A block contains the same number of samples as there are test-conditions involved in the test. A test “session” consists of the same number of blocks as there are talkers involved in the test. Each session is presented to a listening panel of four listeners.
3. Randomizations are constructed such that talker gender is alternated on successive trials resulting in the same talker never being presented on consecutive trials.

Table 2.4.5-1 shows an example randomization for a single listening panel. Each entry in the table is the file name for a sample with the following file-naming convention - ***xxyy.zzz***, where ***xx*** is the talker, ***yy*** is the sample, and ***zzz*** is the test condition.

Table 2.4.5-1 Example Randomization for the Experiments 1 and 3 ACR Test

Panel 1	Blk 1	Blk 2	Blk 3	Blk 4	Blk 5	Blk 6	Blk 7	Blk 8
1	a12f1s6	a18f4s8	a03f2s1	a11f3s6	a06f2s6	a01f2s5	a05f3s8	a09f4s8
2	a11m4s5	a04m2s1	a08m3s7	a22m2s6	a04m4s4	a06m3s5	a15m1s8	a19m2s7
3	a20f1s6	a16f2s1	a21f4s8	a10f2s1	a14f2s2	a03f4s7	a12f2s2	a15f2s4
4	a16m1s1	a20m2s5	a23m2s3	a06m2s4	a17m1s8	a23m4s2	a01m3s3	a20m3s1
5	a05f2s4	a08f2s4	a05f4s8	a19f3s6	a05f1s6	a09f2s4	a13f3s3	a14f1s4
6	a18m3s1	a19m1s1	a06m1s7	a07m3s1	a01m1s6	a12m1s8	a02m4s7	a21m4s3
7	a04f1s4	a23f1s7	a02f1s4	a20f4s5	a07f3s7	a02f3s5	a21f3s5	a01f4s4
8	a09m2s2	a05m3s5	a17m4s1	a13m1s2	a10m2s1	a22m3s8	a18m4s6	a05m4s8
9	a06f3s4	a09f3s7	a18f1s6	a18f2s8	a22f2s4	a08f1s1	a22f4s2	a17f4s8
10	a17m2s5	a11m1s2	a01m4s3	a24m4s6	a20m4s7	a13m2s2	a09m3s4	a04m3s8
11	a23f4s1	a02f4s5	a20f3s2	a17f1s4	a23f3s8	a10f3s2	a11f1s6	a06f1s4
12	a19m4s4	a06m4s8	a15m2s8	a08m4s2	a02m2s5	a07m4s5	a16m2s1	a02m1s4
13	a13f2s6	a24f2s4	a13f4s6	a02f2s7	a16f4s7	a16f1s2	a14f4s7	a07f2s6
14	a08m1s1	a22m4s8	a07m2s7	a14m2s6	a12m4s5	a20m1s4	a10m4s2	a13m4s7
15	a07f4s2	a10f4s6	a10f1s6	a12f4s6	a08f4s2	a11f4s1	a19f1s5	a08f3s7
16	a24m1s2	a03m1s6	a09m4s4	a16m4s7	a18m2s2	a15m4s5	a08m2s3	a10m1s8
17	a21f2s1	a01f3s3	a11f2s4	a01f1s5	a21f1s8	a18f3s1	a03f1s8	a23f2s4
18	a10m3s8	a14m4s7	a22m1s4	a15m3s5	a11m3s4	a04m1s3	a07m1s7	a11m2s5
19	a14f3s6	a15f1s5	a12f3s8	a04f4s1	a24f4s6	a24f1s2	a06f4s6	a16f3s4
20	a03m4s3	a12m2s1	a24m3s4	a23m3s8	a19m3s4	a14m3s4	a24m2s4	a03m2s4
21	a15f4s7	a07f1s7	a04f3s1	a03f3s4	a13f1s2	a17f2s2	a04f2s7	a24f3s4
22	a01m2s6	a13m3s1	a16m3s4	a05m1s5	a09m1s6	a05m2s6	a23m1s5	a18m1s2
23	a22f3s3	a17f3s2	a19f2s7	a09f1s7	a15f3s8	a19f4s1	a20f2s2	a22f1s5
24	a02m3s1	a21m3s4	a14m1s6	a21m1s2	a03m3s4	a21m2s6	a17m3s4	a12m3s3

The randomization lists for each of the eight listening panels for each experiment are provided in /so73/subjctv/exp*/data/play*.lst.

2.4.6 Presentation

Presentation of speech materials for the SO 73 codec listening tests shall be made with one side of high fidelity supra-aural headphones with the other ear uncovered. The speech material delivery system shall meet the requirements of Section 3.4.1.1. The listeners should be seated in a quiet room, with an ambient noise level of 30 dBA or below.

2.4.7 Listeners

The listener sample is intended to represent the population of telephone users with normal hearing acuity. The listeners should be naïve with respect to telephony technology issues; that is, they should not be experts in telephone design, digital voice encoding algorithms, and so on. They should not be trained listeners; that is, they should not have been trained in these or previous listening studies using feedback trials. Age distribution and gender should be nominally balanced across listening panels.

Each listener shall provide data only once for a particular evaluation. A listener may participate in different evaluations, but test sessions performed with the same listener should be at least two months apart so as to reduce the cumulative effects of experience.

2.4.8 Listening Test Procedures

2.4.8.1 ACR Listening Test Procedures – Experiments 1 and 3

The listeners shall listen to each sample and rate the quality of the test sample using a five-point scale, with the points labeled:

5 Excellent

4 Good

3 Fair

2 Poor

1 Bad

Data from 32 listeners shall be used for Experiments 1 and 3, four listeners for each listening panel where each listening panel uses a different randomization. Before starting the test, the listeners should be given instructions for performing the subjective test. An example set of instructions for the ACR are presented in Figure 2.4.8.1-1. The instructions may be modified to allow for variations in laboratory data-gathering apparatus.

This is an experiment to determine the perceived quality of speech over the telephone. You will be listening to a number of recorded speech samples, spoken by several different talkers, and you will be rating how good you think they sound.

Use the single headphone on the ear you normally use for the telephone. On each trial a two-sentence sample will be played. After you have listened to the sample, determine the category from the list below which best describes the overall quality of the sample. Press the numeric key on your keyboard corresponding to your rating for how good or bad that particular passage sounded.

The quality of the speech should be rated according to the scale below:

5 Excellent

4 Good

3 Fair

2 Poor

1 Bad

During the session you will hear samples varying in different aspects of quality. Please take into account your total impression of each sample, rather than concentrating on any particular aspect.

Figure 2.4.8.1-1 Instructions for Listeners

2.4.8.2 P-835 Listening Test Procedures – Experiments 2 and 4

Experiments 2 and 4 use the P.835 test methodology described in [13]. The P.835 methodology is specifically designed to evaluate the quality of speech in background noise. It yields a measure of Signal Quality (SIG), a measure of Background Quality (BAK), and a measure of Overall Quality (OVRL). In general, OVRL scores are highly correlated with MOS but the OVRL score provides greater sensitivity and precision in test conditions involving background noise. While the OVRL score is of most interest here, the SIG and BAK scores also provide valuable diagnostic information. For each trial in a P.835 test, listeners are presented with three sub-samples where each sub-sample is a single sentence (approx. 4 sec. duration) processed through the same test condition. In one of the first two sub-samples listeners rate the *Signal Quality* on a five-point rating scale with the points labeled:

5 Very natural, no distortion
4 Fairly natural, little distortion
3 Somewhat natural, some distortion
2 Fairly unnatural, fairly distorted
1 Very unnatural, very distorted

For the other of the first two sub-samples listeners rate the *Background Quality* on a five-point rating scale with the points labeled:

5	Not noticeable
4	Fairly noticeable
3	Noticeable but not intrusive
2	Fairly conspicuous, somewhat intrusive
1	Very conspicuous, very intrusive

For the third sub-sample listeners rate the *Overall quality* on a five-point rating scale with the points labeled:

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Data from 32 listeners shall be used for Experiments 2 and 4, four listeners for each listening panel where each listening panel uses a different randomization

Before starting the test, the listeners should be given instructions for performing the subjective test. An example set of instructions for the P.835 test are presented below. The instructions may be modified to allow for variations in laboratory data-gathering apparatus.

Instructions for P.835 Speech Rating Experiment

In this speech rating experiment each trial will involve **three** sentences and you will give a rating for **each** sentence.

For the first sentence in each trial you will be asked to attend **only to the speech signal** and rate how natural, or conversely, how degraded, the **speech signal** sounds to you. You will use the rating scale shown in the figure below to register your ratings of the speech signal. Your task will be to choose the numbered phrase from the list below that best describes your opinion of the **SPEECH SIGNAL ALONE** and then enter the corresponding number on your keyboard.

Attending **ONLY to the SPEECH SIGNAL**, select the category which best describes the sample you just heard.

the **SPEECH SIGNAL** in this sample was

5 – VERY NATURAL, NO DEGRADATION
4 – FAIRLY NATURAL, LITTLE DEGRADATION
3 - SOMEWHAT NATURAL, SOMEWHAT DEGRADED
2 – FAIRLY UNNATURAL, FAIRLY DEGRADED
1 – VERY UNNATURAL, VERY DEGRADED

For the second sentence in each trial you will be asked to attend **only to the background** and rate how noticeable, intrusive, and/or conspicuous the **background** sounds to you. You will use the rating scale shown in the figure below to register your ratings of the background. Your task will be to choose the numbered phrase from the list below that best describes your opinion of the **BACKGROUND ALONE** and then enter the corresponding number on your keyboard.

Attending **ONLY to the BACKGROUND**, select the category which best describes the sample you just heard.

the **BACKGROUND** in this sample was

5 – NOT NOTICEABLE
4 – SOMEWHAT NOTICEABLE
3 – NOTICEABLE BUT NOT INTRUSIVE
2 – FAIRLY CONSPICUOUS, SOMEWHAT INTRUSIVE
1 – VERY CONSPICUOUS, VERY INTRUSIVE

For the third and final sentence in each trial you will be asked to attend to the entire sample (both the speech signal and the background) and rate your opinion of the sample for purposes of everyday speech communication.

Select the category which best describes the sample you just heard for purposes of everyday speech communication.

the **OVERALL SPEECH SAMPLE** was

5 – EXCELLENT
4 – GOOD
3 - FAIR
2 – POOR
1 – BAD

2.4.9 Analysis of Results

The response data from the practice blocks shall be discarded. Data sets with missing responses from listeners shall not be used – i.e., a complete set of data is required for 24 listeners, four for each of six listening panels. Responses from the different listening panels for the corresponding test conditions shall be treated as equivalent in the analysis.

2.4.9.1 Basic Results for the SO 73 Listening tests

The votes for each of the test conditions for SO 73 Experiments I and II shall be averaged to produce an associated mean score (***M***) as shown in Equation 2.4.9.1-1 and a Standard Deviation (***SD***) as shown in Equation 2.4.9.1-2, where ***L*** is the number of listeners and ***T*** is the number of talkers involved in the experiment.

$$\overline{M} = \left(\sum_{L} \sum_{T} X_{l,t} \right) \Big/ \left(L \times T \right) \qquad (2.4.9.1\text{-}1)$$

$$SD = \sqrt{\left(\sum_{L} \sum_{T} \left(X_{l,t} - \overline{M} \right)^2 \right) \Big/ \left(L \times T - 1 \right)} \qquad (2.4.9.1\text{-}2)$$

2.4.9.2 Minimum Subjective Requirement for SO 73 Listening Tests

The Terms of Reference for the MPS tests state that the mean score for each of the Test Encoder/Decoder Combinations (E/DC) should be "*not worse than*" the mean score for the Reference E/DC. For most of the test conditions involved in the subjective experiments there are three Test E/DC's (M-T, T-M, and T-T), which means there are three statistical tests against the Reference E/DC (M-M). The three statistical tests are not independent, however. Since they all involve the same ratings for the Reference E/DC, t-tests are not appropriate. The appropriate statistical test for multiple Test conditions against a common Reference condition is Dunnett's Test. A complete description of Dunnett's Test is contained in Appendix B.

The critical value for the Dunnett's test is 2.09 (one-sided test, p<.05, 4 E/DC's, df = 93).

For those test conditions where a single Test E/DC (T-T) is compared against the Reference E/DC (M-M), the appropriate statistical test is Student's t-test[6].

The critical value for the Student's t-test is 1.70 (one-sided test, p<.05, df = 31).

In both the Dunnett's Test and the t-test the MPS test is evaluated by dividing the difference between the mean score for the Test E/DC and the mean score for the Reference ED/C by the Standard Error of the Mean Difference (SE_{MD}) as shown in Equation 2.4.9.2-1. If the resultant Test value is less than

[6] The appropriate t-test is a "matched groups" t-test and the SE_{MD} is based on the differences between individual listener's average ratings, where the average is over talkers. Therefore, the SE_{MD} is based on 32 difference scores, one for each listener (df =31).

the criterion value for the appropriate test (2.09 for Dunnett's Test, 1.70 for the t-test), then the E/DC passes the MPS test.

$$Test = \left(\overline{M}_{Ref} - \overline{M}_{Test}\right) \Big/ SE_{MD} \qquad (2.4.9.2\text{-}1)$$

2.4.10 Expected Results for Reference Conditions

2.4.10.1 Reference Conditions for Experiments 1, 3, and 5

The MNRU conditions have been included to provide a frame of reference for the Experiments 1, 3, and 5. In listening evaluations where test conditions span approximately the same range of quality, the MOS results for similar conditions should be approximately the same. Data from previous studies allows a generalization to be made concerning the expected MOS results for the MNRU reference conditions (see Figure 2.4.10.1-1).

MOS scores obtained for the MNRU conditions in any SO 73 validation test should be compared to those shown in the graph below. Inconsistencies beyond a small shift in the means in either direction or a slight stretching or compression of the scale near the extremes may imply a problem in the execution of the evaluation test. In particular, MOS should be monotonic with MNRU, within the limits of statistical resolution; and the contour of the relation should show a similar slope.

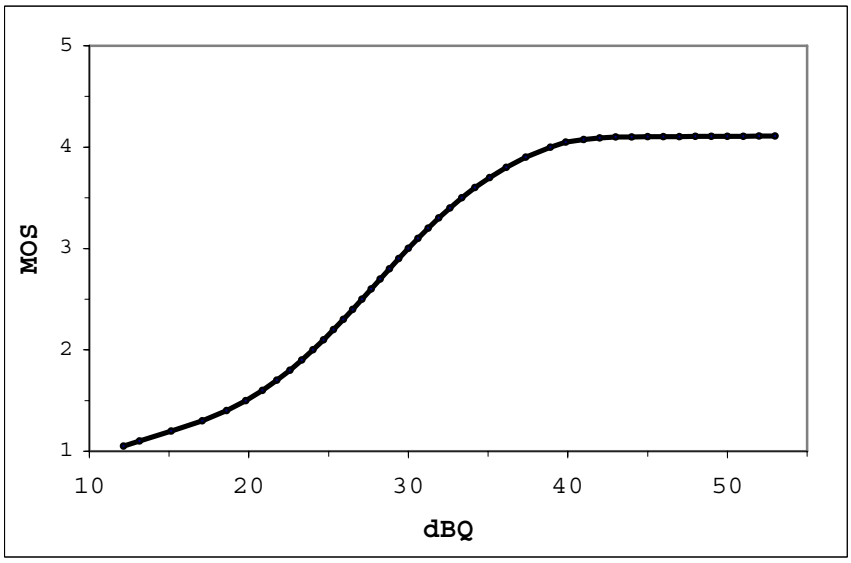


Figure 2.4.10.1-1 Typical Plot of MOS versus MNRU

2.4.10.2 Reference Conditions for Experiments 2 and 4

Reference conditions for P.835 tests are constructed as a combination of SNR and MNRU processing to provide degradation in overall speech quality in two dimensions — signal distortion and background noise intrusiveness. Table 2.4.2.3.2-2 shows the eight reference conditions (b01 – b08)

involved in the P.835 Experiments 2 and 4. In general, results are expected for these reference conditions such that the obtained score profiles are similar to those shown in Figure 2.3.10.1-1.

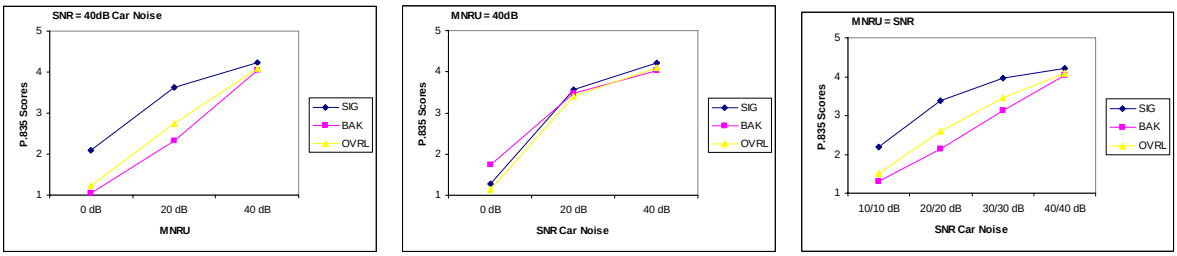


Figure 2.4.10.2-1 Typical P.835 Score Profiles for Reference Conditions

This page intentionally left blank.

3 CODEC STANDARD TEST CONDITIONS

This section describes the conditions, equipment, and the software tools necessary for the performance of the tests of Section 2. The software tools and the speech database associated with Sections 3.1.2 through 3.1.4 (SO 3), 3.2.2 through 3.2.4 (SO 68), 3.3.2 through 3.3.4 (SO 70), or 3.4.2 through 3.4.4 (SO 73) can be found in the Software Distribution associated with this document.

The objective and subjective testing requires that speech data files can be input to the speech encoder and that the output data stream can be saved to a set of files. It is also necessary to input data stream files into the speech decoder and have the output speech data saved to a set of files. This process suggests the use of a computer based data acquisition system to interface to the codec under test. Since the hardware realizations of the speech codec may be quite varied, it is not desirable to precisely define a set of hardware interfaces between such a data acquisition system and the codec. Instead, only a functional description of these interfaces will be defined.

A host computer system is necessary to handle the data files that must be input to the speech encoder and decoder, and to save the resulting output data to files. These data files will contain either sampled speech data or speech codec parameters; hence, all the interfaces are digital. The generic Standard Equipment is shown in Figure 3-1.

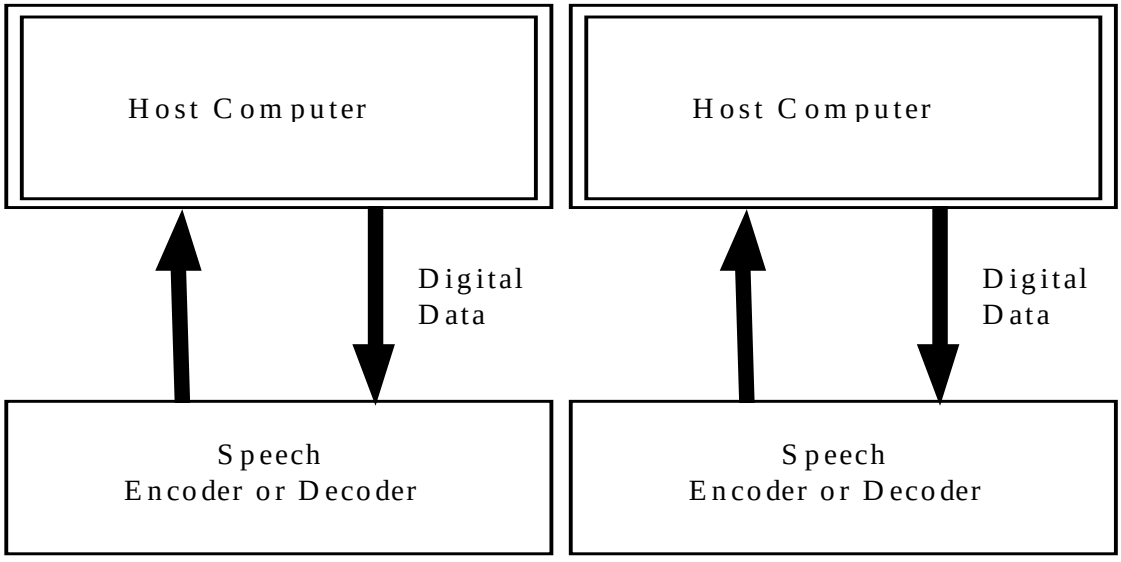


Figure 3-1 Basic Test Equipment

The host computer has access to the data files needed for testing. For encoder testing, the host computer has the source speech data files which it outputs to the speech encoder. The host computer simultaneously saves the speech parameter output data from the encoder. Similarly, for decoder testing, the host computer outputs speech parameters from a disk file and saves the decoder output speech data to a file.

The choice of the host computer and the nature of the interfaces between the host computer and the speech codec are not subject to standardization. It is expected that the host computer would be some type of personal computer or workstation with suitable interfaces and adequate disk storage. The

interfaces may be serial or parallel and will be determined by the interfaces available on the particular hardware realization of the speech codec.

Figure 3-2 shows a generic block diagram of the audio path for the subjective test using four listeners per session. The audio path is shown as a solid line; the data paths for experimental control are shown as broken lines. This figure is for explanatory purposes and does not prescribe a specific implementation.

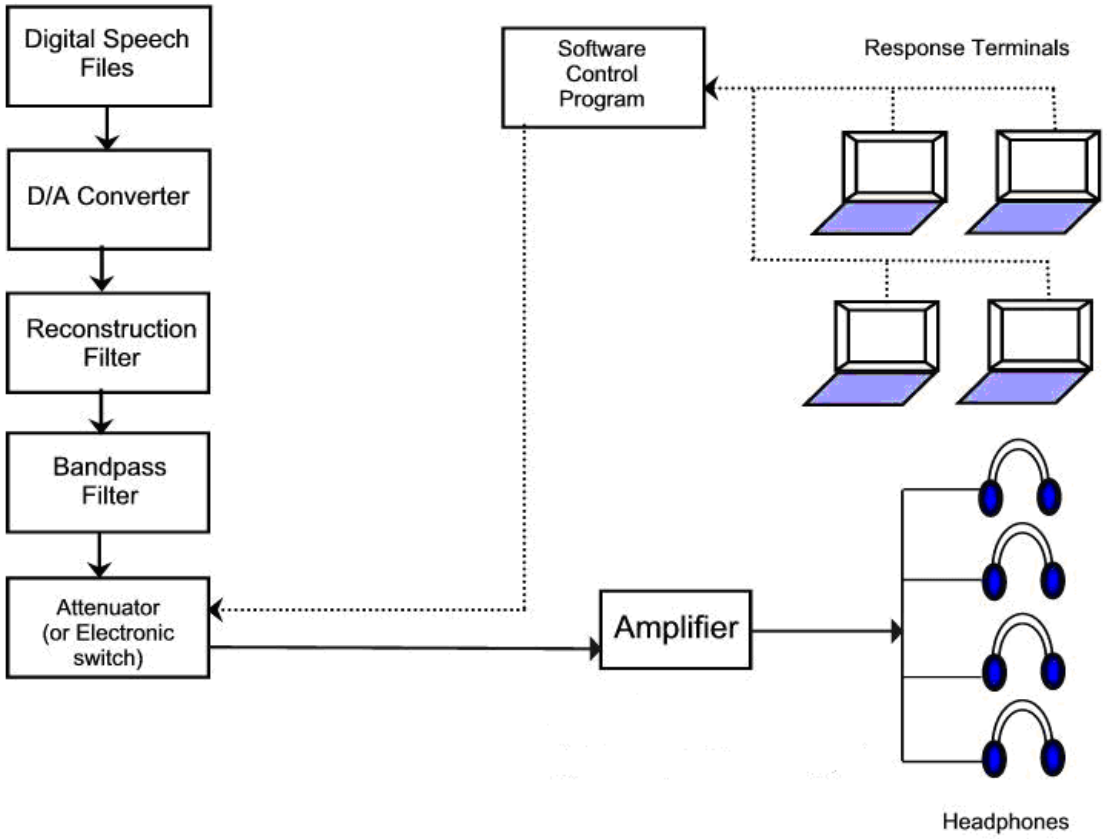


Figure 3-2 Subjective Testing Equipment Configuration

For the purposes of this standard, speech codecs under test are not required to provide performance across ranges of temperature, humidity or other typical physical environmental variables.

3.1 Specific Standard Test Conditions for SO 3

3.1.1 Audio Path and Calibration for SO 3

3.1.1.1 Audio Path

The audio path must meet the following requirements for electro-acoustic performance measured between the output of the D/A converter and the output of the headphone:

1. Frequency response shall be flat to within ±2 dB between 200 Hz and 3400 Hz, and below 200 Hz the response shall roll off at a minimum of 12 dB per octave. Equalization may be used in the audio path to achieve this. A suitable reconstruction filter shall be used for playback.
2. Total harmonic distortion shall be less than 1% for signals between 100 Hz and 4000 Hz.
3. Noise over the audio path shall be less than 35 dBA measured at the ear reference plane of the headphone.
4. Signal shall be delivered to the headphone on the listener's preferred telephone ear. No signal shall be delivered to the other headphone.

3.1.1.2 Calibration

The audio circuit shall deliver an average sound level of the stimuli to the listener at -16 dBPa (78 dB SPL) at the ear reference plan. This level was chosen because it is equivalent to the level delivered by a nominal ROLR handset driven by the average signal level on the PSTN network. This level may be calibrated using a suitable artificial ear with circum-aural headphone adapter and microphone. A test file with a reference signal is included with the source speech database for the purpose of calibration. The file cos1004_.290 is located in the directory /so3/cal of the companion software. The calibration file contains a -22 dB 1004 Hz reference signal. The audio circuit shall be calibrated so that the test signal has a level of -16 dBPa at the ear reference plane, while maintaining compliance with Section 3.1.1.1.

3.1.2 Standard Software Test Tools for SO 3

This section describes a set of software tools useful for performing the tests specified in Section 2.1. Where possible, code is written in C-code [19] and has been developed and compiled using the GNU GCC[7] C-language compiler and software maintenance utilities. The tools have been verified under various representative operating systems on a number of different hardware platforms. The 3GPP2-supplied tools are all located in the /so3/tools directory in the associated Software Distribution, and can be built using the GNU make utility, using static libraries, and no special optimizations, by copying the contents of the /so3/tools directory to a new directory on a writeable disk and typing “make all” in that directory. A GCC compatible makefile has been provided for this purpose in the /so3/tools directory. The makefile creates the executables avg_rate (.exe), l_mu_l (.exe), and sv56 (.exe) in the /so3/tools/bin directory. This makefile may need to be modified to conform to the user’s hardware platform.

---

[7] The GNU-C compiler (GCC) and software development tools including documentation are available without charge from the Free Software Foundation. They can be contacted at:

Free Software Foundation	Voice: +1-617-542-5942
59 Temple Place - Suite 330	Fax: +1-617-542-2652
Boston, MA 02111-1307, USA	gnu@gnu.org

or on the World Wide Web at http://www.fsf.org

Those non-3GPP2 supplied tools (l_mu_l (.exe), and sv56 (.exe)), available in C-code form from [6] and [6a], and compiled using GCC, are identified, and are to be used supplementary to those available on the Software Distribution.

The program descriptions that follow all use the convention of enclosing optional command line arguments in angle brackets (<>).

3.1.2.1 Average Data Rate Determination Utility - avg_rate.c

This utility program is used to determine the average data rate at which a test codec encodes a set of benchmark speech files. The source code, avg_rate.c, is a 3GPP2-supplied tool and is located in the /so3/tools/avg_rate directory of the associated Software Distribution. The input to the program is a list of packet file names, where each packet file referred to in the list conforms to the format described in Section 3.1.3.3. The output of the program is, for each file referred to by the input file list: The file name, the number of packets contained in the file, and the average data rate calculated as described in Section 2.1.1.1. The average data rate utility is intended to be used on the packet files created by the test codec in response to the average rate benchmark files referred to in Section 2.1.1.1 and located in the /so3/objctv directory of the associated Software Distribution. The program is invoked as follows:

avg_rate filename_1 <filename_2> <filename_3> ... <filename_n>

3.1.2.2 Scaling speech files - sv56.c

This program is used to scale each sample in a linearly quantized speech file by a factor that renders the file's root mean square (RMS) level equal to a user-specified value. The program is intended to be used on the test codec's speech output files to ensure that their RMS level is consistent with the requirements of Section 2.1.2.3 of this document. The source code, sv56.c, is available from [6] and [6a]. The inputs to the program are the (optional) desired RMS value in dB, the input speech file name and the (optional) output speech file name. The outputs are the initial (prior to scaling) maximum sample, RMS and average (DC) values in the speech file, the final (after scaling) maximum, RMS, and DC values in the output file, the number of samples that were clipped, the scale factor applied, and an output speech file appropriately scaled. If no target RMS value is specified, the program calculates and prints the initial statistics mentioned above and copies the input file to the output file unmodified. The program is invoked as follows:

sv56 Desired-RMS-Level File_In File_Out [Sample Rate (Resolution)]

Note: The desired level specified for sv56 differs by 3dB from the value required for this specification. For example, in order to adjust speech files -22dB in accordance with this specification, the calling sequence is:

sv56 -25 File_In File_Out

3.1.2.3 μ-Law Companding - l_mu_l.c

This program applies μ-Law companding to the sample values in a linearly quantized speech file according to [7]. The source code, l_mu_l.c, is available from [6] and [6a]. The input to the program is the speech file to be companded. The output is the companded speech file. Both files are linearly quantized speech files in accordance with Section 3.1.3.3 of this document. The program is invoked as follows:

l_mu_l input_filename output_filename

3.1.3 Master Codec for SO 3

This section describes the C simulation of the speech codec specified by [1]. The master codec C simulation used for verifying the performance of a non-bit-exact EVRC implementation shall be the floating-point master C simulation included in the associated Software Distribution [1a].

3.1.3.1 Compiling the Master Codec Simulation

The source code for floating-point C simulation has been written in ANSI C and compiled using the GNU GCC C compiler and make utility. Refer to Section 3.1.2 for information regarding obtaining GCC make and relevant documentation.

A GCC compatible makefile has been included in [1a]. Typing "make" in the appropriate directory will compile and link the code and create the executable file called EvrcFlt (evrcflt.exe on Win32 systems). The included makefile may require some user modification for a particular hardware platform and/or operating system.

3.1.3.2 Running the Master Codec Simulation

The EVRC executable files use command line arguments to receive all information regarding input and output files and various parameters used during execution.

Executing EvrcFlt with no command line arguments will display a brief description of the required and optional command line arguments. The options are described below:

-i infn (required)	Specifies the name of the input speech file, or the name of the input packet file if only decoding is being performed (see the -d option below).
-o outf (required)	Specifies the name of the output speech file, or the name of the output packet file if only encoding is being performed (see the -e option below).
-d	Instructs the simulation to perform only the decoding function. The input file must contain packets of compressed data.
-e	Instructs the simulation to perform only the encoding function. The output file will contain packets of compressed data.
	If neither the -d or the -e option is invoked, the coder performs both the encoding and decoding functions by default.
-h max	Sets the maximum allowable data rate to max, where max is element of 4, 3, 1, using the codes specified in the first column of Table 3.1.3.3-1.
-l min	Sets the minimum allowable data rate to min, where min is element of 4,3,1, using the codes specified in the first column of Table 3.1.3.3-1.
	If neither the -h nor -l option is invoked, the coder allows the data rate to vary between Rate 1 and Rate 1/8.
	In addition, if max ≠ min, the data rate varies between max and min using the same rate decision algorithm, where the data rate is set to max if the selected data rate is >= max, and the data rate is set to min if the

selected data rate is <= min. See the select_rate() routine in the file ratedec.c for more information.

-p flag — If flag is set to 0, the post-filter is disabled. If the flag is set to 1, the post-filter is enabled. If the -p option is not invoked, the post-filter is enabled during decoding.

-n flag — If flag is set to 0, noise suppression is disabled. If the flag is set to 1, noise suppression is enabled. If the -n option is not invoked, noise suppression is enabled during encoding.

3.1.3.3 File Formats

Files of speech contain 2's complement 16-bit samples with the least significant byte first. The packet file contains twelve 16-bit words with the low byte ordered first followed by the high byte.

The first word in the packet contains the data rate while the remaining 11 words contain the encoded speech data packed in accordance with the tables specified in [1]. The packet file value for each data rate is shown in Table 3.1.3.3-1.

Table 3.1.3.3-1 Packet File Structure From Master Codec/Channel Error Model

Value in Packet File	Rate	Data Bits per Frame
4 = 0x0004	1	171
3 = 0x0003	1/2	80
1 = 0x0001	1/8	16
0 = 0x0000	Blank	0
15 = 0x000f	Full Rate Probable	171
14 = 0x000e	Erasure	0

Unused bits are set to 0. For example, in a Rate 1/8 frame, the packet file will contain the word 0x0100 (byte-swapped 0x0001) followed by one 16-bit word containing the 16 data bits for the frame (in byte-swapped form), followed by ten 16-bit words containing all zero bits.

3.1.3.4 Verifying Proper Operation of the Master Codec

Files are provided for the purpose of verifying the fixed-point codec executable.

Three files, mstr_ref.pcm, mstr_ref.pkt, and mstr_ref.dec, are included in the directory /master/test to provide a means for verifying proper operation of the master codec software. The file mstr_ref.pcm is an unprocessed speech file. The file mstr_ref.pkt is a packet file that was obtained by running

```
EvrcFlt -i mstr_ref.pcm -o mstr_ref.pkt -e
```

The file mstr_ref.dec is a decoded speech file that was obtained by running

```
EvrcFlt -i mstr_ref.pkt -o mstr_ref.dec -d
```

Once EvrcFlt is compiled, verification files should be processed as follows:

```
EvrcFlt -i mstr_ref.pcm -o verify.pkt -e
```

EvrcFlt -i verify.pkt -o verify.dec -d

If the output files mstr_ref.pkt and mstr_ref.dec exactly match the verify.pkt and the verify.dec, respectively, then verification of the master codec's operation is complete.

Because of differences in the way that floating-point arithmetic is done in different computing environments, it will not always be true that the floating-point master C simulation will produce identical output in response to the same input when compiled and run on different compiler/hardware platforms, even though the simulation is operating correctly. In the event that the exact match described in the preceding paragraph is not obtained, it is recommended that the user verify that the version of GCC used is version 2.7.2 or later.

3.1.4 Fixed-Point Bit-Exact Codec for SO 3

This section describes the C simulation of the speech codec specified by [1]. The speech codec C simulation is based on finite precision, fixed-point arithmetic operations and is required to be used as a reference codec to verify the performance of a bit-exact EVRC implementation of the fixed-point C simulation of a test codec. The bit-exact EVRC codec, along with the appropriate test vectors to verify the bit-exactness performance, are included in the associated Software Distribution.

There are two options for compiling the fixed point EVRC simulation. One option uses the 31-bit long multiply DSP math library and the other uses the 32-bit library. A parallel set of bit-exact test vectors is provided so that a CODEC may qualify as bit-exact using either library.

3.1.4.1 Fixed-Point Codec Program Files

This section describes the C program files which are provided in the directory /so3/simul/fixed in the companion software. All of the files needed to compile, run, and verify the fixed-point codec are located in the directory /so3/simul/fixed.

3.1.4.2 Compiling the Fixed-Point Codec Simulation

The source code for the fixed-point codec simulation has been written in ANSI C and can be compiled using any general purpose compiler such as the GNU GCC C compiler and make utility. Refer to Section 3.3 for information regarding obtaining GCC, make, and relevant documentation.

Two GCC compatible makefiles have been included in the /so3/simul/fixed/code and /so3/simul/fixed/dspmath directory. All of the files contained on the associated Software Distribution under the directory /fixed should be copied onto a writable disk, making sure to preserve the directory structure. Typing "make" in the */dspmath directory first, followed by typing "make" in the directory */code will compile and link the code and create the executable file called EvrcFix (evrcfix.exe on Win32 systems), which will be placed in the */bin directory. The included makefiles may require some user modification for a particular hardware platform and/or operating system.

There exists two options for compiling the fixed point EVRC simulation. One option uses the 31-bit long multiply DSP math library and the other uses the 32-bit library. A parallel set of bit-exact test vectors is provided so that a CODEC may qualify as bit-exact using either library.

By default, the DSP math library compiles the 32-bit long multiply routines. In order to compile with the 31-bit long multiply routines, the following lines in /so3/simul/fixed/dspmath/makefile must be commented/uncommented:

Change from 32-bit library:

#Uncomment the following line to use alternate double precision multiplies

#CCAUXFLAGS=-DUSE_ALT_DP31

#& comment the following line out

CCAUXFLAGS=

to 31-bit library:

#Uncomment the following line to use alternate double precision multiplies

CCAUXFLAGS=-DUSE_ALT_DP31

#& comment the following line out

#CCAUXFLAGS=

3.1.4.3 Running the Fixed-Point Codec Simulation

The EVRC executable files use command line arguments to receive all information regarding input and output files and various parameters used during execution.

Executing EvrcFix with no command line arguments will display a brief description of the required and optional command line arguments. The options are described below:

-i infn (required)	Specifies the name of the input speech file, or the name of the input packet file if only decoding is being performed (see the -d option below).
-o outf (required)	Specifies the name of the output speech file, or the name of the output packet file if only encoding is being performed (see the -e option below).
-d	Instructs the simulation to perform only the decoding function. The input file must contain packets of compressed data.
-e	Instructs the simulation to perform only the encoding function. The output file will contain packets of compressed data.
	If neither the -d or the -e option is invoked, the coder performs both the encoding and decoding functions by default.
-f max	Sets the maximum number of frames to be processed.
-h max	Sets the maximum allowable data rate to max, where max is element of 4, 3, 1, using the codes specified in the first column of Table 3.1.3.3-1.
-l min	Sets the minimum allowable data rate to min, where min is element of 4,3,1, using the codes specified in the first column of Table 3.1.3.3-1.
	If neither the -h nor -l option is invoked, the coder allows the data rate to vary between Rate 1 and Rate 1/8.
	In addition, if max ≠ min, the data rate varies between max and min using the same rate decision algorithm, where the data rate is set to max if the selected data rate is >= max, and the data rate is set to min if the selected data rate is <= min. See the select_rate() routine in the file ratedec.c for more information.

-p flag	If flag is set to 0, the post-filter is disabled. If the flag is set to 1, the post-filter is enabled. If the -p option is not invoked, the post-filter is enabled during decoding.
-n flag	If flag is set to 0, noise suppression is disabled. If the flag is set to 1, noise suppression is enabled. If the -n option is not invoked, noise suppression is enabled during encoding.

3.1.4.4 File Formats

Files of speech contain 2's complement 16-bit samples with the least significant byte first. The packet file contains twelve 16-bit words with the low byte ordered first followed by the high byte.

The first word in the packet contains the data rate while the remaining 11 words contain the encoded speech data packed in accordance with the tables specified in [1]. The packet file value for each data rate is shown in Table 3.1.3.3-1. Unused bits are set to 0. For example, in a Rate 1/8 frame, the packet file will contain the word 0x0100 (byte-swapped 0x0001) followed by one 16-bit word containing the 16 data bits for the frame (in byte-swapped form), followed by ten 16-bit words containing all zero bits.

3.1.4.5 Verifying Proper Operation of the Fixed-Point Codec

Files are provided for the purpose of verifying the fixed-point codec executable.

The files /so3/simul/fixed/test/source/*.pcm contain the original, unprocessed speech files. The files in /so3/simul/fixed/test/fixed32 contain the encoded packet files and the decoded speech files generated by the 32-bit long multiply DSP library. Likewise, files in /so3/simul/fixed/test/fixed31 were processed with the 31-bit DSP library. The processed files have the following naming convention. The encoded packet have the extension *.pkt and are generated by running

EvrcFix -i *.pcm -o *.pkt -e

the decoded speech files, *.dec, are generated by running

EvrcFix -i *.pkt -o *.dec -d

If the output files *.pkt and *.dec exactly match verify_*.pkt and verify_*.dec, respectively, then verification of the operation of the fixed-point codec's operation is complete.

3.1.4.6 Verifying Bit-Exact Performance of the Fixed-Point Test Codec

Files in the /so3/testvec directory are provided for the purpose of qualifying a test codec as bit-exact. The files in the /so3/testvec/* directories are 16 bit PCM binary files in PC format (LSB,MSB) and obey the following file extension naming convention:

source speech:	*.pcm
encoder output:	*.pkt
decoder output:	*.dec

The /so3/testvec directory is divided into 3 subdirectories: /so3/testvec/source, /so3/testvec/fixed31, and /so3/testvec/fixed32.

The /so3/testvec/source directory contains input source files and includes original speech files as well as packet files injected with frame erasures. The /so3/testvec/fixed31 (/so3/testvec/fixed32) directory contains files processed with the 31-bit (32-bit) DSP library. The files in these directories are the reference files for bit-exact compliance. A test codec is bit-exact if it can reproduce all of the reference files in either the /so3/testvec/fixed32 directory or the /so3/testvec/fixed31 directory.

3.1.4.6.1 Description of Bit-Exact Source Files

The following source files are designed to exercise the majority of the bitstream slots.

vec_01.pcm	15dB babble	7 females, 7 males
vec_02.pcm	10dB car	7 females, 7 males
vec_03.pcm	flat clean	7 females, 7 males
vec_04.pcm	15dB street	7 females, 7 males
vec_05.pcm	high level	4 females, 4 males
vec_06.pcm	low level	4 females, 4 males
vec_07.pcm	irs clean	4 females, 4 males
vec_08.pcm	flat clean	4 females, 4 males
vec_09.pcm	10dB car	4 females, 4 males
vec_10.pcm	15dB babble	4 females, 4 males
vec_11.pcm	12dB street	4 females, 4 males
vec_12.pcm	mixed noise one-sided conversation	
vec_13.pcm	mixed noise one-sided conversation	

The following source files are designed to exercise the RCELP algorithm.

NOTE: These files must be processed in full-rate only mode (only rate 4 allowed).

shiftr.pcm	Frequency-sweep
shiftl.pcm	Frequency-sweep

The following source files are recordings of one-sided conversations at different input levels and are designed to test the rate determination algorithm.

rda_test.pcm

rda_mod.pcm

rda_high.pcm

rda_low.pcm

The following source files are encoded packets which have been corrupted with frame erasure at different rates. They are designed to exercise the decoder's frame error handling.

vec_07_1.pkt	Encoded packet w/ 1% FER
vec_07_2.pkt	Encoded packet w/ 2% FER
vec_07_3.pkt	Encoded packet w/ 3% FER
vec_08_1.pkt	Encoded packet w/ 1% FER
vec_08_2.pkt	Encoded packet w/ 2% FER
vec_08_3.pkt	Encoded packet w/ 3% FER
vec_10_1.pkt	Encoded packet w/ 1% FER
vec_10_2.pkt	Encoded packet w/ 2% FER
vec_10_3.pkt	Encoded packet w/ 3% FER

3.1.4.6.2 Instructions for Processing Bit-Exact Test Vectors

The following table is a list of source files to be processed in DEFAULT MODE (rates 1,3,4 allowed) and the names of the corresponding reference files. The files are to be processed as follows:

Encode: EvrcFix -e -i file.pcm -o file.pkt

Decode: EvrcFix -d -i file.pkt -o file.dec

Table 3.1.4.6.2-1 Source and Bit-exact Default Mode Test Vector Files

PCM Source File	Encoded Packet File	Decoded Speech File
rda_high.pcm	rda_high.pkt	rda_high.dec
rda_low.pcm	rda_low.pkt	rda_low.dec
rda_mod.pcm	rda_mod.pkt	rda_mod.dec
rda_test.pcm	rda_test.pkt	rda_test.dec
vec_01.pcm	vec_01.pkt	vec_01.dec
vec_02.pcm	vec_02.pkt	vec_02.dec
vec_03.pcm	vec_03.pkt	vec_03.dec
vec_04.pcm	vec_04.pkt	vec_04.dec
vec_05.pcm	vec_05.pkt	vec_05.dec
vec_06.pcm	vec_06.pkt	vec_06.dec

PCM Source File	Encoded Packet File	Decoded Speech File
vec_07.pcm	vec_07.pkt	vec_07.dec
vec_08.pcm	vec_08.pkt	vec_08.dec
vec_09.pcm	vec_09.pkt	vec_09.dec
vec_10.pcm	vec_10.pkt	vec_10.dec
vec_11.pcm	vec_11.pkt	vec_11.dec
vec_12.pcm	vec_12.pkt	vec_12.dec
vec_13.pcm	vec_13.pkt	vec_13.dec

The following table is a list of source files to be processed in Rate-1/2 Maximum (rates 1,3 allowed) and the names of the corresponding reference files. The files are to be processed as follows:

Encode: EvrcFix -e -h 3 -i file.pcm -o file_h.pkt

Decode: EvrcFix -d -i file_h.pkt -o file_h.dec

Table 3.1.4.6.2-2 Source and Bit-exact Rate-1/2 Max Test Vector Files

PCM Source File	Encoded Packet File	Decoded Speech File
vec_05.pcm	vec_05_h.pkt	vec_05_h.dec
vec_06.pcm	vec_06_h.pkt	vec_06_h.dec
vec_08.pcm	vec_08_h.pkt	vec_08_h.dec

The following table is a list of source files to be processed in FULL RATE ONLY MODE (only rate 4 allowed) and the names of the corresponding reference files. The files are to be processed as follows:

Encode: EvrcFix -e -l 4 -i file.pcm -o file.pkt

Decode: EvrcFix -d -i file.pkt -o file.dec

Table 3.1.4.6.2-3 Source and Bit-exact Full Rate Only Test Vector Files

PCM Source File	Encoded Packet File	Decoded Speech File
shiftl.pcm	shiftl.pkt	shiftl.dec
shiftr.pcm	shiftr.pkt	shiftr.dec

The following table is a list of source packet files to be decoded and the names of the corresponding reference files. Note that it is not necessary to reproduce the source packet files, only the decoded speech files. The files are to be processed as follows:

Decode: EvrcFix -d -i file.pkt -o file.dec

Table 3.1.4.6.2-4 Decoder Output Test Vector Files

Packet Source File	Decoded Speech File
vec_07_1.pkt	vec_07_1.dec
vec_07_2.pkt	vec_07_2.dec
vec_07_3.pkt	vec_07_3.dec
vec_08_1.pkt	vec_08_1.dec
vec_08_2.pkt	vec_08_2.dec
vec_08_3.pkt	vec_08_3.dec
vec_10_1.pkt	vec_10_1.dec
vec_10_2.pkt	vec_10_2.dec
vec_10_3.pkt	vec_10_3.dec

3.2 Specific Standard Test Conditions for SO 68

3.2.1 Audio Path and Calibration for SO 68

3.2.1.1 Audio Path

The audio path must meet the following requirements for electro-acoustic performance measured between the output of the D/A converter and the output of the headphone:

1. Frequency response shall be flat to within ±2 dB between 200 Hz and 3400 Hz, and below 200 Hz the response shall roll off at a minimum of 12 dB per octave. Equalization may be used in the audio path to achieve this. A suitable reconstruction filter shall be used for playback.
2. Total harmonic distortion shall be less than 1% for signals between 100 Hz and 4000 Hz.
3. Noise over the audio path shall be less than 35 dBA measured at the ear reference plane of the headphone.
4. Signal shall be delivered to the headphone on the listener's preferred telephone-listening ear, and the other ear shall be uncovered. No signal shall be delivered to the other headphone.

3.2.1.2 Calibration

The audio circuit shall deliver an average sound level of the stimuli to the listener at -15 dBPa (79 dB SPL) at the ear reference plan. This level was chosen because it is equivalent to the level delivered by a nominal ROLR handset driven by the average signal level on the PSTN network. This level may be calibrated using a suitable artificial ear with circum-aural headphone adapter and microphone. A test file with a reference signal is included with the source speech database for the purpose of calibration. The file cos1004_.290 is located in the directory /so68/cal of the companion software. The calibration file contains a -22 dB 1004 Hz reference signal. The audio circuit shall be

calibrated so that the test signal has a level of -15 dBPa at the ear reference plane, while maintaining compliance with Section 3.2.1.1.

3.2.2 Standard Software Test Tools for SO 68

This section describes a set of software tools useful for performing the MPS tests. The code has been developed and compiled using the GNU g++[8] compiler and software maintenance utilities. The tools have been verified under various representative operating systems on a number of different hardware platforms. The 3GPP2 supplied tools are all located in the /so68/tools directory in the associated Software Distribution, and can be built using the GNU g++ compiler.

Other software tools such as `scaldemo, actlev, filter,` and `astrip` are available in [6].

3.2.2.1 Channel Model Utilities – `fersig27(.exe)`

This utility program provides

a) the ability to introduce Frame Erasure channel impairment.

b) the ability to verify use of half-rate or lesser frame rate during dim-and-burst and packet level signaling

c) the ability to measure the Average Data Rate from an encoded packet file

A log output of `fersig27` provides detail on the ADR performance of the preceding encoder. In these applications, the utility is invoked as in following examples for 3% FER, and 1% signaling:

```
fersig27 -c EVRC-B -e fer_3%.bin infile outfile

fersig27 -c EVRC-B -s dim_1%.bin -e fer_3%.bin infile outfile
```

3.2.2.2 Channel Error and Signaling Masks

These binary Frame Error Rate and Signaling masks (source level and packet level) (1 byte of either 0 or 1 per frame) are used with the `fersig27` channel-impairment and inter-working simulation functions for the various conditions:

```
fer_3%.bin

dim_1%.bin

dim_1%_pls.bin
```

---

[8] The GNU-C compiler (G++) and software development tools including documentation are available without charge from the Free Software Foundation. They can be contacted at:

Free Software Foundation	Voice: +1-617-542-5942
59 Temple Place - Suite 330	Fax: +1-617-542-2652
Boston, MA 02111-1307, USA	gnu@gnu.org

or on the World Wide Web at http://www.fsf.org

3.2.2.3 EVRC-B Interworking Function (IWF)

The software “EvrcB_iwf.cc” can be compiled to yield a simulation utility `EvrcB_iwf` with usage defined as:

```
EvrcB_iwf –s signaling_mask_file –i encoded_packet_file –o dimmed_packet_file
```

where `EvrcB_iwf` converts full-rate frames in the input “encoded_packet_file” to half-rate frames at packet-level (that is using a simple scaling down of the packet instead of a complicated transcoding method).

3.2.3 Master Codec for SO 68

This section describes the C simulation of the speech codec specified by [1]. The master codec C simulation used for verifying the performance of a non-bit-exact EVRC-B implementation shall be the floating-point master C simulation included in the associated Software Distribution [1a].

3.2.3.1 Compiling the Master Codec Simulation

The source code for floating-point simulation can be compiled using the GNU G++ compiler and make utility.

A G++ compatible makefile has been included in the appropriate sub-directory in [1a]. Typing "make" this directory will compile and link the code and create the executable file called EvrcB (EvrcB.exe on Win32 systems), which will be placed in the same directory. The included makefile may require some user modification for a particular hardware platform and/or operating system.

3.2.3.2 Running the Master Codec Simulation

The EVRC-B floating point executable (EvrcB) files use command line arguments to receive all information regarding input and output files and various parameters used during execution.

Executing “EvrcB” with no command line arguments will display a brief description of the required and optional command line arguments. The options are described below:

-i infn (required)	Specifies the name of the input speech file, or the name of the input packet file if only decoding is being performed (see the -d option below).
-o outf (required)	Specifies the name of the output speech file, or the name of the output packet file if only encoding is being performed (see the -e option below).
-d	Instructs the simulation to perform only the decoding function. The input file must contain packets of compressed data.
-e	Instructs the simulation to perform only the encoding function. The output file will contain packets of compressed data.
	If neither the -d or the -e option is invoked, the coder performs both the encoding and decoding functions by default.
-M max	Sets the maximum allowable data rate to max, where max is element of 4, 3, 2, 1, using the codes specified in the first column of Table 3.2.3.3-1.

-m min — Sets the minimum allowable data rate to min, where min is element of 4, 3, 2, 1, using the codes specified in the first column of Table 3.2.3.3-1.

If neither the -M nor -m option is invoked, the coder allows the data rate to vary between Rate 1 and Rate 1/8.

-W <target_active_speech_channel_adr>

Specifies the target active speech channel average data rate in **kbps** that the EVRC-B encoder should target. For example -W 7.5 for 7.5 kbps.

3.2.3.3 File Formats for SO 68

Files of speech contain 2's complement 16-bit samples with the least significant byte first. The packet file contains twelve 16-bit words with the low byte ordered first followed by the high byte.

The first word in the packet contains the data rate while the remaining 11 words contain the encoded speech data packed in accordance with the tables specified in [1]. The packet file value for each data rate is shown in Table 3.2.3.3-1.

Table 3.2.3.3-1 Packet File Structure From Master Codec/Channel Error Model

Value in Packet File	Rate	Data Bits per Frame
4 = 0x0004	1	171
3 = 0x0003	1/2	80
1 = 0x0001	1/8	16
0 = 0x0000	Blank	0
14 = 0x000e	Erasure	0

Unused bits are set to 0. For example, in a Rate 1/8 frame, the packet file will contain the word 0x0100 (byte-swapped 0x0001) followed by one 16-bit word containing the 16 data bits for the frame (in byte-swapped form), followed by ten 16-bit words containing all zero bits.

3.2.4 Fixed-Point Bit-Exact Codec for SO 68

This section describes the C simulation of the speech codec specified by [1]. The speech codec C simulation is based on finite precision, fixed-point arithmetic operations and is recommended to be used as a reference codec to verify the performance of a bit-exact EVRC-B implementation of the fixed-point C simulation of a test codec. The bit-exact EVRC-B codec, along with the appropriate test vectors to verify the bit-exactness performance, are included in the associated Software Distribution.

3.2.4.1 Fixed-Point Codec Program Files

This section describes the C program files which are provided in the associated software distribution for this document. All of the files needed to compile, run, and verify the fixed-point codec are located in the directory /so68/EVRCB_FX.

3.2.4.2 Compiling the Fixed-Point Codec Simulation

The source code for the fixed-point codec simulation has been written in C++ and can be compiled using any general purpose compiler such as the GNU G++ compiler and make utility. Refer to Section 3.3 for information regarding obtaining GCC, make, and relevant documentation.

Two GCC compatible makefiles have been included in the /so68/EVRCB_FX/build directory. Typing “make” in the /build directory will compile and link the code and create the executable file called EvrcB_fx (EvrcB_fx.exe on Win32 systems), which will be placed in the /build directory. The included makefiles may require some user modification for a particular hardware platform and/or operating system.

3.2.4.3 Running the Fixed-Point Codec Simulation

The EVRC-B executable files use command line arguments to receive all information regarding input and output files and various parameters used during execution.

Executing EvrcB_fx with no command line arguments will display a brief description of the required and optional command line arguments. The options are described below:

-i infn (required)	Specifies the name of the input speech file, or the name of the input packet file if only decoding is being performed (see the -d option below).
-o outf (required)	Specifies the name of the output speech file, or the name of the output packet file if only encoding is being performed (see the -e option below).
-d	Instructs the simulation to perform only the decoding function. The input file must contain packets of compressed data.
-e	Instructs the simulation to perform only the encoding function. The output file will contain packets of compressed data.
	If neither the -d or the -e option is invoked, the coder performs both the encoding and decoding functions by default.
-M max	Sets the maximum allowable data rate to max, where max is element of 4, 3, 2, 1, using the codes specified in the first column of Table 3.2.3.3-1.
-m min	Sets the minimum allowable data rate to min, where min is element of 4, 3, 2, 1, using the codes specified in the first column of Table 3.2.3.3-1.
	If neither the -M nor -m option is invoked, the coder allows the data rate to vary between Rate 1 and Rate 1/8.
	In addition, if max ≠ min, the data rate varies between max and min using the same rate decision algorithm, where the data rate is set to max if the selected data rate is >= max, and the data rate is set to min if the selected data rate is <= min.
-W <target_active_speech_channel_adr>	
	Specifies the target active speech channel average data rate in **bps** that the EVRC-B encoder should target. For example -W 7500 for 7.5 kbps

3.2.4.4 File Formats

Files of speech contain 2's complement 16-bit samples with the least significant byte first. The packet file contains twelve 16-bit words with the low byte ordered first followed by the high byte.

The first word in the packet contains the data rate while the remaining 11 words contain the encoded speech data packed in accordance with the tables specified in [1]. The packet file value for each data rate is shown in Table 3.2.3.3-1. Unused bits are set to 0. For example, in a Rate 1/8 frame, the packet file will contain the word 0x0100 (byte-swapped 0x0001) followed by one 16-bit word containing the 16 data bits for the frame (in byte-swapped form), followed by ten 16-bit words containing all zero bits.

3.2.4.5 Verifying Bit-Exact Performance of the Fixed-Point Test Codec

Files in the `/so68/testvec/` directory are provided for the purpose of qualifying a test codec as bit-exact, and conform to the file-naming convention described in Section 2.2.4:

The `/so68/testvec` directory is divided into 2 subdirectories: `/so68/testvec/source`, and `/so68/testvec/fixed`.

The `/so68/testvec/source` directory contains input source files as well as packet files injected with frame erasures. The `/so68/testvec/fixed` directory contains files processed with the EVRC-B fixed point reference software. The files in these directories are the reference files for bit-exact compliance. A test codec is bit-exact if it can reproduce all of the reference files in the `/so68/testvec/fixed` directory exactly. The outputs of the encoder and decoder of the test codec are to be obtained for the conditions given below in Table 3.2.4.5-1 and Table 3.2.4.5-2 The processing steps for these conditions are illustrated in Section 6.

Table 3.2.4.5-1 SO 68 Encoder Bit-exact Test Conditions

Input File	Operating Point	Condition	Reference packet files for bit-exact compliance
src.s22	EVRC-B 9.3 kbps	Nominal, -22 dB	9_3.p22
src.s22	EVRC-B 5.8 kbps	Nominal, -22 dB	5_8.p22
src.s22	EVRC-B 4.8 kbps	Nominal, -22 dB	4_8.p22
src.s32	EVRC-B 9.3 kbps	Low, -32 dB, 1% d&b	9_3.p32
src.s32	EVRC-B 5.8 kbps	Low, -32 dB, 1% d&b	5_8.p32
src.s12	EVRC-B 9.3 kbps	High, -12 dB	9_3.p12
src.s12	EVRC-B 5.8 kbps	High, -12 dB	5_8.p12
src.c15	EVRC-B 9.3 kbps	Nominal, -22 dB, 15 dB carnoise	9_3.pc
src.c15	EVRC-B 5.8 kbps	Nominal, -22 dB, 15 dB carnoise	5_8.pc
src.b20	EVRC-B 9.3 kbps	Nominal, -22 dB, 20 dB babble	9_3.po
src.b20	EVRC-B 5.8 kbps	Nominal, -22 dB, 20 dB babble	5_8.po
src.s15	EVRC-B 9.3 kbps	Nominal, -22 dB, 15 dB street	9_3.ps
src.s15	EVRC-B 5.8 kbps	Nominal, -22 dB, 15 dB street	5_8.ps

Note: 9.3 kbps mode is generated using anchor operating point 0 and 5.8 kbps mode is generated using anchor operating point 2.

Table 3.2.4.5-2 SO 68 Decoder Bit-exact Test Conditions

Input Packet File	Operating Point	Condition	Reference output speech files for bit-exact compliance
9_3.p22	EVRC-B 9.3 kbps	Nominal, -22 dB	9_3.o22
5_8.p22	EVRC-B 5.8 kbps	Nominal, -22 dB	5_8.o22
4_8.p22	EVRC-B 4.8 kbps	Nominal, -22 dB	4_8.o22
9_3.p32	EVRC-B 9.3 kbps	Low, -32 dB, 1% d&b, 1% pls	9_3.o32
5_8.p32	EVRC-B 5.8 kbps	Low, -32 dB, 1% d&b, 1% pls	5_8.o32
9_3.p12	EVRC-B 9.3 kbps	High, -12 dB	9_3.o12
5_8.p12	EVRC-B 5.8 kbps	High, -12 dB	5_8.o12
9_3.pc	EVRC-B 9.3 kbps	Nominal, -22 dB, 15 dB carnoise	9_3.oc
5_8.pc	EVRC-B 5.8 kbps	Nominal, -22 dB, 15 dB carnoise	5_8.oc
9_3.po	EVRC-B 9.3 kbps	Nominal, -22 dB, 20 dB babble	9_3.ob
5_8.po	EVRC-B 5.8 kbps	Nominal, -22 dB, 20 dB babble	5_8.ob
9_3.ps	EVRC-B 9.3 kbps	Nominal, -22 dB, 15 dB street	9_3.os
5_8.ps	EVRC-B 5.8 kbps	Nominal, -22 dB, 15 dB street	5_8.os

3.3 Specific Standard Test Conditions for SO 70

3.3.1 Audio Path and Calibration for SO 70

3.3.1.1 Audio Path

The audio path for wideband test conditions (Experiments 1 and 2) must meet the following requirements for electro-acoustic performance measured between the output of the D/A converter and the output of the headphone:

1. Frequency response shall be flat to within ±2 dB between 50 Hz and 7000 Hz, and below 50 Hz the response shall roll off at a minimum of 12 dB per octave. Equalization may be used in the audio path to achieve this. A suitable reconstruction filter shall be used for playback.
2. Total harmonic distortion shall be less than 1% for signals between 50 Hz and 8000 Hz.
3. Noise over the audio path shall be less than 35 dBA measured at the ear reference plane of the headphone.
4. Signal shall be delivered to the headphone on the listener's preferred telephone-listening ear, and the other ear shall be uncovered. No signal shall be delivered to the other headphone.

The audio path for narrowband test conditions (Experiments 3, 4, 5, and 6) must meet the following requirements for electro-acoustic performance measured between the output of the D/A converter and the output of the headphone:

1. Frequency response shall be flat to within ±2 dB between 200 Hz and 3400 Hz, and below 200 Hz the response shall roll off at a minimum of 12 dB per octave. Equalization may be used in the audio path to achieve this. A suitable reconstruction filter shall be used for playback.
2. Total harmonic distortion shall be less than 1% for signals between 100 Hz and 4000 Hz.
3. Noise over the audio path shall be less than 35 dBA measured at the ear reference plane of the headphone.
4. Signal shall be delivered to the headphone on the listener's preferred telephone-listening ear, and the other ear shall be uncovered. No signal shall be delivered to the other headphone.

3.3.1.2 Calibration

The audio circuit shall deliver an average sound level of the stimuli to the listener at -18 dBPa (76 dB SPL) at the ear reference plan. This level was chosen because it is equivalent to the level delivered by a nominal ROLR handset driven by the average signal level on the PSTN network. This level may be calibrated using a suitable artificial ear with circum-aural headphone adapter and microphone. A test file with a reference signal is included with the source speech database for the purpose of calibration. The file cal_1004.16k is located in the directory /so70/cal of the companion software. The calibration file contains a -22 dB 1004 Hz reference signal. The audio circuit shall be calibrated so that the test signal has a level of -15 dBPa at the ear reference plane, while maintaining compliance with Section 3.3.1.1.

3.3.2 Software Test Tools for SO 70

This section describes a set of software tools useful for performing the MPS tests. The code has been developed and compiled using the GNU g++ compiler and software maintenance utilities. The tools have been verified under various representative operating systems on a number of different hardware platforms. The 3GPP2 supplied tools are all located in the /so70/tools directory in the associated Software Distribution, and can be built using the GNU g++ compiler.

Other software tools such as `scaldemo`, `actlev`, `filter`, and `astrip` are available in [6].

3.3.2.1 Channel Model Utilities – `fersig28(.exe)`

This utility program provides

a) the ability to introduce Frame Erasure channel impairment.

b) the ability to verify use of half-rate or lesser frame rate during dim-and-burst and packet level signaling

c) the ability to measure the Average Data Rate from an encoded packet file

A log output of `fersig28` provides detail on the ADR performance of the preceding encoder. In these applications, the utility is invoked as in following examples for 3% FER, and 1% signaling:

```
fersig28 -c EVRC-WB -e fer_3%.bin infile outfile

fersig28 -c EVRC-WB -s dim_1%.bin -e fer_3%.bin infile outfile
```

3.3.2.2 Channel Error and Signaling Masks

These binary Frame Error Rate and Signaling masks (source level and packet level) (1 byte of either 0 or 1 per frame) are used with the `fersig28` channel-impairment and inter-working simulation functions for the various conditions:

```
fer_3%.bin

dim_1%.bin

dim_1%_pls.bin
```

3.3.2.3 EVRC-WB Interworking Function (IWF)

The software “Evrc_wb_iwf.cc” can be compiled to yield a simulation utility `Evrc_wb_iwf` with usage defined as:

```
Evrc_wb_iwf –s signaling_mask_file –i encoded_packet_file –o dimmed_packet_file
```

where `Evrc_wb_iwf` converts full-rate frames in the input “encoded_packet_file” to half-rate frames at packet-level (that is using a simple scaling down of the packet instead of a complicated transcoding method).

3.3.2.4 P.341 Tx Filter

The software utility “p341_tx.c” can be compiled to yield a Tx filtering utility `p341_tx` with usage defined as:

```
p341_tx input-file-name output-file-name
```

where p341_tx is the 3GPP2 Tx filter compliant to ITU-T P.341. Figure 3.3.2.4-1 shows the frequency response of “p341_tx” filter. Also shown in this figure is the response of the ITU-T P.341 STL-2000 filter implementation, as well as the transmit masks for the ITU-T P.341/P.311 and the wideband transmit response from Table 9 in the 3GPP electro-acoustics specification [21]. From this figure, it can be seen that the STL-2000 filter response (in red) does not meet the frequency response of the 3GPP electro-acoustics specification, while the p341_tx filter response (in green) meets both the P.341/P.311 masks as well as the 3GPP electro-acoustics specification mask.

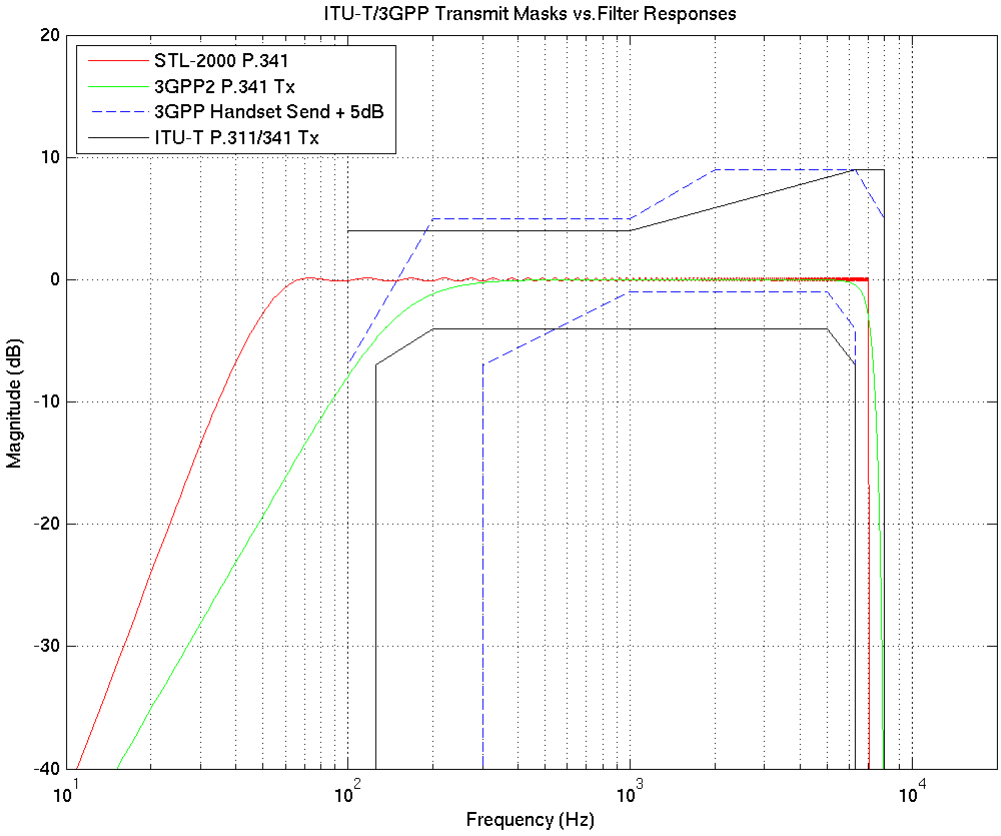


Figure 3.3.2.4-1 SO 70 ITU-T P.311/P.341 Transmit Mask and Filter responses

3.3.3 Master Codec for SO 70

This section describes the C simulation of the speech codec specified by [1]. The master codec C simulation used for verifying the performance of a non-bit-exact EVRC-WB implementation shall be the floating-point master C simulation included in the associated Software Distribution [1a].

3.3.3.1 Compiling the Master Codec Simulation

The source code for floating-point simulation can be compiled using the GNU G++ compiler and make utility.

A G++ compatible makefile has been included in the appropriate sub-directory in [1a]. Typing "make" this directory will compile and link the code and create the executable file called Evrc_wb (Evrc_wb.exe on Win32 systems), which will be placed in the same directory. The included makefile may require some user modification for a particular hardware platform and/or operating system.

3.3.3.2 Running the Master Codec Simulation

The EVRC-WB floating point executable (Evrc_wb) files use command line arguments to receive all information regarding input and output files and various parameters used during execution.

Executing “Evrc_wb” with no command line arguments will display a brief description of the required and optional command line arguments. The options are described below:

-i infn (required)	Specifies the name of the input speech file, or the name of the input packet file if only decoding is being performed (see the -d option below).
-o outf (required)	Specifies the name of the output speech file, or the name of the output packet file if only encoding is being performed (see the -e option below).
-d	Instructs the simulation to perform only the decoding function. The input file must contain packets of compressed data.
-e	Instructs the simulation to perform only the encoding function. The output file will contain packets of compressed data.
	If neither the -d or the -e option is invoked, the coder performs both the encoding and decoding functions by default.
-M max	Sets the maximum allowable data rate to max, where max is element of 4, 3, 1, using the codes specified in the first column of Table 3.3.3.3-1.
-m min	Sets the minimum allowable data rate to min, where min is element of 4, 3, 1, using the codes specified in the first column of Table 3.1.3.3-1.
	If neither the -M nor -m option is invoked, the coder allows the data rate to vary between Rate 1 and Rate 1/8.

3.3.3.3 File Formats for SO 70

Files of speech contain 2’s complement 16-bit samples with the least significant byte first. The packet file contains twelve 16-bit words with the low byte ordered first followed by the high byte.

The first word in the packet contains the data rate while the remaining 11 words contain the encoded speech data packed in accordance with the tables specified in [1]. The packet file value for each data rate is shown in Table 3.1.3.3-1.

Table 3.3.3.3-1 Packet File Structure From Master Codec/Channel Error Model

Value in Packet File	Rate	Data Bits per Frame
4 = 0x0004	1	171

Value in Packet File	Rate	Data Bits per Frame
3 = 0x0003	1/2	80
1 = 0x0001	1/8	16
0 = 0x0000	Blank	0
14 = 0x000e	Erasure	0

Unused bits are set to 0. For example, in a Rate 1/8 frame, the packet file will contain the word 0x0100 (byte-swapped 0x0001) followed by one 16-bit word containing the 16 data bits for the frame (in byte-swapped form), followed by ten 16-bit words containing all zero bits.

3.3.4 Fixed-Point Bit-Exact Codec for SO 70

This section describes the C simulation of the speech codec specified by [1]. The speech codec C simulation is based on finite precision, fixed-point arithmetic operations and is recommended to be used as a reference codec to verify the performance of a bit-exact EVRC-WB implementation of the fixed-point C simulation of a test codec. The bit-exact EVRC-WB codec, along with the appropriate test vectors to verify the bit-exactness performance, are included in the associated Software Distribution.

3.3.4.1 Fixed-Point Codec Program Files

This section describes the C program files which are provided in the associated software distribution for this document.

3.3.4.2 Compiling the Fixed-Point Codec Simulation

The source code for the fixed-point codec simulation has been written in C++ and can be compiled using any general purpose compiler such as the GNU G++ compiler and make utility.

Two GCC compatible makefiles have been included in the /build directory. Typing “make” in the /build directory will compile and link the code and create the executable file called Evrc_wb_fx (Evrc_wb_fx.exe on Win32 systems), which will be placed in the /build directory. The included makefiles may require some user modification for a particular hardware platform and/or operating system.

3.3.4.3 Running the Fixed-Point Codec Simulation

The EVRC-WB executable files use command line arguments to receive all information regarding input and output files and various parameters used during execution.

Executing Evrc_wb_fx with no command line arguments will display a brief description of the required and optional command line arguments. The options are described below:

-i infn (required) Specifies the name of the input speech file, or the name of the input packet file if only decoding is being performed (see the -d option below).

-o outf (required) Specifies the name of the output speech file, or the name of the output packet file if only encoding is being performed (see the -e option below).

-d	Instructs the simulation to perform only the decoding function. The input file must contain packets of compressed data.
-e	Instructs the simulation to perform only the encoding function. The output file will contain packets of compressed data.
	If neither the -d or the -e option is invoked, the coder performs both the encoding and decoding functions by default.
-M max	Sets the maximum allowable data rate to max, where max is element of 4, 3, 1, using the codes specified in the first column of Table 3.1.3.3-1.
-m min	Sets the minimum allowable data rate to min, where min is element of 4, 3, 1, using the codes specified in the first column of Table 3.1.3.3-1.
	If neither the -M nor -m option is invoked, the coder allows the data rate to vary between Rate 1 and Rate 1/8.
	In addition, if max ≠ min, the data rate varies between max and min using the same rate decision algorithm, where the data rate is set to max if the selected data rate is >= max, and the data rate is set to min if the selected data rate is <= min.

3.3.4.4 File Formats

Files of speech contain 2's complement 16-bit samples with the least significant byte first. The packet file contains twelve 16-bit words with the low byte ordered first followed by the high byte.

The first word in the packet contains the data rate while the remaining 11 words contain the encoded speech data packed in accordance with the tables specified in [1]. The packet file value for each data rate is shown in Table 3.1.3.3-1. Unused bits are set to 0. For example, in a Rate 1/8 frame, the packet file will contain the word 0x0100 (byte-swapped 0x0001) followed by one 16-bit word containing the 16 data bits for the frame (in byte-swapped form), followed by ten 16-bit words containing all zero bits.

3.3.4.5 Verifying Bit-Exact Performance of the Fixed-Point Test Codec

This section outlines the methodology of verifying whether a Fixed-point Test codec is bit-exact to the Fixed point reference software. The purpose of this testing is to evaluate the bit-exactness of the test codec under a variety of conditions which may occur. To accomplish this, suites of test vectors have been designed to test for bit-exactness of the Test Codec under a variety of conditions depending on a number of parameters. These conditions include channel impairments, audio background noise, and different input levels.

Figure 3.3.4.5-1 illustrates a decision tree to arrive at the suite of test-vectors that are needed to demonstrate Minimum Performance Spec compliance through bit-exactness of a Test implementation of SO 70 for different profiles of equipments that support SO 70.

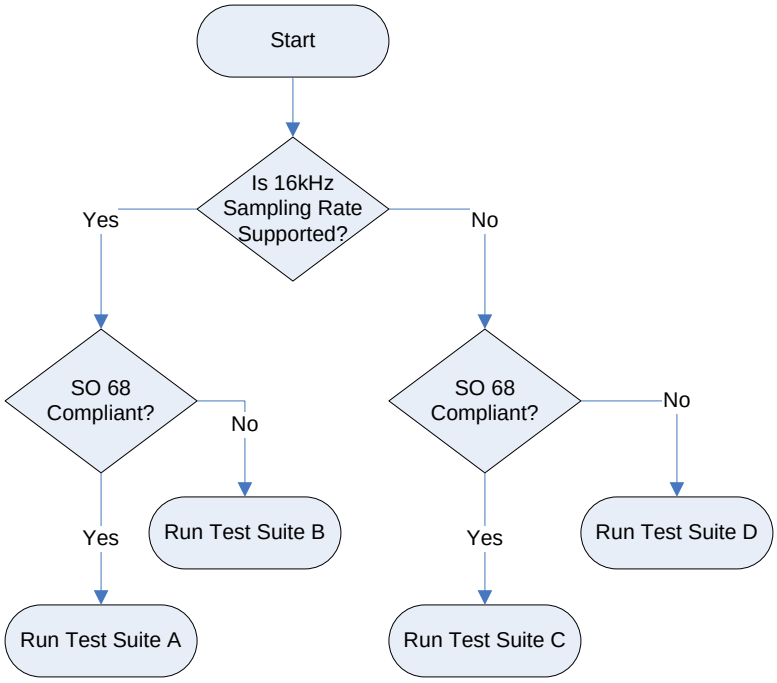


Figure 3.3.4.5-1 SO 70 Fixed-point bit-exact test suite decision flowchart

An implementation may support SO 70 only for 8 kHz sample rate input/output (for example, a Base-station transcoder or a Media Gateway). An implementation may support SO 70 for both 16 kHz and 8 kHz sample rate (for example, a mobile station that supports wideband electro-acoustics).

Further, the implementation supporting SO 70 might already have demonstrated compliance to SO 68 Minimum Performance Spec. This means that such an equipment has also demonstrated the Minimum Performance requirements for RATE_REDUC operating points 4 and 7 of SO 70 (which exactly correspond to the RATE_REDUC operating points 0 and 7 of SO 68).

Therefore, the main parameters in the decision tree are a) 16 kHz support in the implementation b) SO 68 compliance of the test implementation.

Depending on the implementation profile of the Device under test, one of 4 possible Test Suites are to be used to demonstrate SO 70 compliance. These 4 test suites named Test suites A, B, C, D, and the individual input test vectors comprising the Test suites are highlighted in Table 3.3.4.5-1.

Table 3.3.4.5-1 Test Suites of input test vectors for SO 70 compliance

Test Suites	Directory containing input test vectors	Notes
A	/so70/testvec/source/suiteA	Mobile application already supporting SO 68 compliance
B	/so70/testvec/source/suiteB	Mobile application NOT already supporting SO 68 compliance
C	/so70/testvec/source/suiteC	Infra/MGW application already supporting SO 68 compliance
D	/so70/testvec/source/suiteD	Infra/MGW application NOT already supporting SO 68 compliance

Files in the /so70/testvec/ directory are provided for the purpose of qualifying a test codec as bit-exact, and conform to the file-naming convention described in Section 2.2.4:

The /so70/testvec directory is divided into 2 subdirectories: /so70/testvec/source, and /so70/testvec/fixed.

The /so70/testvec/source directory contains input source files as well as packet files injected with frame erasures. The /so70/testvec/fixed directory contains files processed with the EVRC-WB fixed point reference software. The files in these directories are the reference files for bit-exact compliance. A test codec is bit-exact if it can reproduce all of the reference files in the /so70/testvec/fixed directory exactly. The outputs of the encoder and decoder of the test codec are to be obtained for the conditions given below in Table 3.3.4.5-2 -- Table 3.3.4.5-9. The processing steps for these conditions are illustrated in Section 6.

Table 3.3.4.5-2 SO 70 Encoder Suite A Bit-exact Test Conditions

Input File	Operating Point	Condition	Reference packet files for bit-exact compliance
src.s22	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB	evrc_wb_op0.p22
src.s12	EVRC-WB, operating point 0, 16 kHz sampling	High, -12 dB	evrc_wb_op0.p12
src.s32	EVRC-WB, operating point 0, 16 kHz sampling	Low, -32 dB, 1% d&b	evrc_wb_op0.dim_1%.p32
src.c10	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 10 dB car noise	evrc_wb_op0.pc1
src.c20	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB car noise	evrc_wb_op0.pc2
src.s15	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_wb_op0.ps
src.b20	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_wb_op0.pb

Table 3.3.4.5-3 SO 70 Suite A Decoder Bit-exact Test Conditions

Input Packet File	Operating Point	Condition	Reference output speech files for bit-exact compliance
evrc_wb_op0.fer_3%.p22	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 3% FER	evrc_wb_op0.fer_3%.o22
evrc_wb_op0.fer_1%.pls_1%.p22	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 3% FER	evrc_wb_op0. fer_1%.pls_1%.o22
evrc_wb_op0. p12	EVRC-WB, operating point 0, 16 kHz sampling	High, -12 dB,	evrc_wb_op0.o12
evrc_wb_op0.dim_1%.p32	EVRC-WB, operating point 0, 16 kHz sampling	Low, -32 dB, 1% d&B	evrc_wb_op0.dim_1%.o32
evrc_wb_op0.pc1	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 10 dB car noise	evrc_wb_op0.oc1
evrc_wb_op0._fer_3%.pc2	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB car noise, fer_3%	evrc_wb_op0.fer_3%.oc2
evrc_wb_op0.ps	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_wb_op0.os
evrc_wb_op0.pb	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_wb_op0.ob
evrc_wb_op0.fer_3%.pm	EVRC-WB, operating point 0, 16 kHz sampling	Generic audio signal, fer_3%	evrc_wb_op0.fer_3%.om

Table 3.3.4.5-4 SO 70 Encoder Suite B Bit-exact Test Conditions

Input File	Operating Point	Condition	Reference packet files for bit-exact compliance
src.s22	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB	evrc_wb_op0.p22
src.s12	EVRC-WB, operating point 0, 16 kHz sampling	High, -12 dB	evrc_wb_op0.p12
src.s32	EVRC-WB, operating point 0, 16 kHz sampling	Low, -32 dB, 1% d&b	evrc_wb_op0.p32
src.c10	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 10 dB car noise	evrc_wb_op0.pc1

Input File	Operating Point	Condition	Reference packet files for bit-exact compliance
src.c20	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB car noise	evrc_wb_op0.pc2
src.s15	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_wb_op0.ps
src.b20	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_wb_op0.pb
src.s22.8k	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB	evrc_wb_op4.p22
src.s12.8k	EVRC-WB, operating point 4, 8 kHz sampling	High, -12 dB	evrc_wb_op4.p12
src.s32.8k	EVRC-WB, operating point 4, 8 kHz sampling	Low, -32 dB	evrc_wb_op4.p32
src.s22.8k	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 1% d&b	evrc_wb_op4.dim_1%.p22
src.s22.8k	EVRC-WB, operating point 7, 8 kHz sampling	Nominal, -22 dB	evrc_wb_op7.p22
src.c15.8k	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_wb_op4.pc
src.s15.8k	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_wb_op4.ps
src.b20.8k	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_wb_op4.pb
src.c15.8k	EVRC-WB, operating point 7, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_wb_op7.pc

Table 3.3.4.5-5 SO 70 Suite B Decoder Bit-exact Test Conditions

Input Packet File	Operating Point	Condition	Reference output speech files for bit-exact compliance
evrc_wb_op0.fer_3%.p22	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 3% FER	evrc_wb_op0.fer_3%.o22
evrc_wb_op0.fer_1%.pls_1%.p22	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 3% FER	evrc_wb_op0. fer_1%.pls_1%.o22
evrc_wb_op0.p12	EVRC-WB, operating point 0, 16 kHz sampling	High, -12 dB	evrc_wb_op0.o12
evrc_wb_op0.dim_1%.p32	EVRC-WB, operating point 0, 16 kHz sampling	Low, -32 dB, 1% d&B	evrc_wb_op0.dim_1%.o32

Input Packet File	Operating Point	Condition	Reference output speech files for bit-exact compliance
evrc_wb_op0..pc1	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 10 dB car noise	evrc_wb_op0.oc1
evrc_wb_op0._fer_3%.pc2	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB car noise, fer_3%	evrc_wb_op0.fer_3%.oc2
evrc_wb_op0.ps	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_wb_op0.os
evrc_wb_op0.po	EVRC-WB, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_wb_op0.ob
evrc_wb_op0.fer_3%.pm	EVRC-WB, operating point 0, 16 kHz sampling	Generic audio signal, fer_3%	evrc_wb_op0.fer_3%.om
evrc_wb_op4.fer_3%.p22	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, FER 3%	evrc_wb_op4.fer_3%.o22.8k
evrc_wb_op4.p12	EVRC-WB, operating point 4, 8 kHz sampling	High, -12 dB	evrc_wb_op4.o12.8k
evrc_wb_op4.p32	EVRC-WB, operating point 4, 8 kHz sampling	Low, -32 dB	evrc_wb_op4.o32.8k
evrc_wb_op7.p22	EVRC-WB, operating point 7, 8 kHz sampling	Nominal, -22 dB	evrc_wb_op7.o22.8k
evrc_wb_op4.dim_1%.pls_1%.p22	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 1% d&b, 1% pls	evrc_wb_op4.dim_1%.pls_1% .o22.8k
evrc_wb_op4.pc	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_wb_op4.oc.8k
evrc_wb_op7.pc	EVRC-WB, operating point 7, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_wb_op7.oc.8k
evrc_wb_op4.ps	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_wb_op4.os.8k
evrc_wb_op4.fer_2%.pb	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 15 dB babble noise	evrc_wb_op4.fer_2%.ob.8k

Table 3.3.4.5-6 SO 70 Encoder Suite C Bit-exact Test Conditions

Input File	Operating Point	Condition	Reference packet files for bit-exact compliance
		No need encoder tests if SO 68 already proven	

Table 3.3.4.5-7 SO 70 Suite C Decoder Bit-exact Test Conditions

Input Packet File	Operating Point	Condition	Reference output speech files for bit-exact compliance
evrc_wb_op0.dim_1%.fer_2%.p22	EVRC-WB, operating point 0, 8 kHz sampling	Nominal, -22 dB, 1% d&b, 2% FER	evrc_wb_op0.dim_1%.fer_2%.o22.8k
evrc_wb_op0.pls_1%.p22	EVRC-WB, operating point 0, 8 kHz sampling	Nominal, -22 dB, 1% pls	evrc_wb_op0.pls_1%.o22.8k
evrc_wb_op0.p12	EVRC-WB, operating point 0, 8 kHz sampling	High, -12 dB	evrc_wb_op0.o12.8k
evrc_wb_op0.p32	EVRC-WB, operating point 0, 8 kHz sampling	Low, -32 dB	evrc_wb_op0.o32.8k
evrc_wb_op0.dim_2%.pc	EVRC-WB, operating point 0, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise, 2% d&b	evrc_wb_op0.dim_2%.oc.8k
evrc_wb_op0.pls_1%.pc	EVRC-WB, operating point 0, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise, 1% pls	evrc_wb_op0.pls_1%.oc.8k
evrc_wb_op0.ps	EVRC-WB, operating point 0, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_wb_op0.os.8k
evrc_wb_op0.fer_2%.pb	EVRC-WB, operating point 0, 8 kHz sampling	Nominal, -22 dB, 20 dB babble noise, 2% FER	evrc_wb_op0.fer_2%.ob.8k
evrc_wb_op0.fer_3%.pm	EVRC-WB, operating point 0, 8 kHz sampling	Generic audio signal, fer_3%	evrc_wb_op0.fer_3%.om.8k

Table 3.3.4.5-8 SO 70 Encoder Suite D Bit-exact Test Conditions

Input File	Operating Point	Condition	Reference packet files for bit-exact compliance
src.s22.8k	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB	evrc_wb_op4.p22
src.s12.8k	EVRC-WB, operating point 4, 8 kHz sampling	High, -12 dB	evrc_wb_op4.p12
src.s32.8k	EVRC-WB, operating point 4, 8 kHz sampling	Low, -32 dB	evrc_wb_op4.p32
src.s22.8k	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 1% d&b	evrc_wb_op4.dim_1%.p22
src.s22.8k	EVRC-WB, operating point 7, 8 kHz sampling	Nominal, -22 dB	evrc_wb_op7.p22
src.c15.8k	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_wb_op4.pc
src.s15.8k	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_wb_op4.ps
src.b20.8k	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_wb_op4.pb
src.c15.8k	EVRC-WB, operating point 7, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_wb_op7.pc

Table 3.3.4.5-9 SO 70 Suite D Decoder Bit-exact Test Conditions

Input Packet File	Operating Point	Condition	Reference output speech files for bit-exact compliance
evrc_wb_op0.fer_3%.p22	EVRC-WB, operating point 0, 8 kHz sampling	Nominal, -22 dB, 3% FER	evrc_wb_op0.fer_3%.o22.8k
evrc_wb_op0.p12	EVRC-WB, operating point 0, 8 kHz sampling	High, -12 dB	evrc_wb_op0.o12.8k
evrc_wb_op0.p32	EVRC-WB, operating point 0, 8 kHz sampling	Low, -32 dB	evrc_wb_op0.o32.8k
evrc_wb_op0.dim_1%.pls_1%.p22	EVRC-WB, operating point 0, 8 kHz sampling	Nominal, -22 dB, 1% d&b, 1% pls	evrc_wb_op0.dim_1%.pls_1% .o22.8k
evrc_wb_op0.pc	EVRC-WB, operating point 0, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_wb_op0.oc.8k

Input Packet File	Operating Point	Condition	Reference output speech files for bit-exact compliance
evrc_wb_op0.ps	EVRC-WB, operating point 0, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_wb_op0.os.8k
evrc_wb_op0.fer_2%.pb	EVRC-WB, operating point 0, 8 kHz sampling	Nominal, -22 dB, 20 dB babble noise, 2% FER	evrc_wb_op0.fer_2%.ob.8k
evrc_wb_op0.fer_3%.pm	EVRC-WB, operating point 0, 8 kHz sampling	Generic audio signal, fer_3%	evrc_wb_op0.fer_3%.om.8k
evrc_wb_op4.fer_3%.p22	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, FER 3%	evrc_wb_op4.fer_3%.o22.8k
evrc_wb_op4.p12	EVRC-WB, operating point 4, 8 kHz sampling	High, -12 dB	evrc_wb_op4.o12.8k
evrc_wb_op4.p32	EVRC-WB, operating point 4, 8 kHz sampling	Low, -32 dB	evrc_wb_op4.o32.8k
evrc_wb_op7.p22	EVRC-WB, operating point 7, 8 kHz sampling	Nominal, -22 dB	evrc_wb_op7.o22.8k
evrc_wb_op4.dim_1%.pls_1%.p22	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 1% d&b, 1% pls	evrc_wb_op4.dim_1%.pls_1% .o22.8k
evrc_wb_op4.pc	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_wb_op4.oc.8k
evrc_wb_op7.pc	EVRC-WB, operating point 7, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_wb_op7.oc.8k
evrc_wb_op4.ps	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_wb_op4.os.8k
evrc_wb_op4.fer_2%.pb	EVRC-WB, operating point 4, 8 kHz sampling	Nominal, -22 dB, 15 dB babble noise	evrc_wb_op4.fer_2%.pb.8k

1

3.4 Specific Standard Test Conditions for SO 73

3.4.1 Audio Path and Calibration for SO 73

3.4.1.1 Audio Path

The audio path for wideband test conditions (Experiments 1 and 2) must meet the following requirements for electro-acoustic performance measured between the output of the D/A converter and the output of the headphone:

1. Frequency response shall be flat to within ±2 dB between 50 Hz and 7000 Hz, and below 50 Hz the response shall roll off at a minimum of 12 dB per octave. Equalization may be used in the audio path to achieve this. A suitable reconstruction filter shall be used for playback.
2. Total harmonic distortion shall be less than 1% for signals between 50 Hz and 8000 Hz.
3. Noise over the audio path shall be less than 35 dBA measured at the ear reference plane of the headphone.
4. Signal shall be delivered to the headphone on the listener's preferred telephone-listening ear, and the other ear shall be uncovered. No signal shall be delivered to the other headphone.

The audio path for narrowband test conditions (Experiments 3 and 4) must meet the following requirements for electro-acoustic performance measured between the output of the D/A converter and the output of the headphone:

1. Frequency response shall be flat to within ±2 dB between 200 Hz and 3400 Hz, and below 200 Hz the response shall roll off at a minimum of 12 dB per octave. Equalization may be used in the audio path to achieve this. A suitable reconstruction filter shall be used for playback.
2. Total harmonic distortion shall be less than 1% for signals between 100 Hz and 4000 Hz.
3. Noise over the audio path shall be less than 35 dBA measured at the ear reference plane of the headphone.
4. Signal shall be delivered to the headphone on the listener's preferred telephone-listening ear, and the other ear shall be uncovered. No signal shall be delivered to the other headphone.

3.4.1.2 Calibration

The audio circuit shall deliver an average sound level of the stimuli to the listener at -18 dBPa (76 dB SPL) at the ear reference plan. This level was chosen because it is equivalent to the level delivered by a nominal ROLR handset driven by the average signal level on the PSTN network. This level may be calibrated using a suitable artificial ear with circum-aural headphone adapter and microphone. A test file with a reference signal is included with the source speech database for the purpose of calibration. The file cal_1004.16k is located in the directory /so73/cal of the companion software. The calibration file contains a -22 dB 1004 Hz reference signal. The audio circuit shall be calibrated so that the test signal has a level of -15 dBPa at the ear reference plane, while maintaining compliance with Section 3.4.1.1.

3.4.2 Software Test Tools for SO 73

This section describes a set of software tools useful for performing the MPS tests. The code has been developed and compiled using the GNU g++ compiler and software maintenance utilities. The tools have been verified under various representative operating systems on a number of different hardware platforms. The 3GPP2 supplied tools are all located in the /so73/tools directory in the associated Software Distribution, and can be built using the GNU g++ compiler.

Other software tools such as `scaldemo`, `actlev`, `filter`, and `astrip` are available in [6].

3.4.2.1 Channel Model Utilities – `fersig29(.exe)`

This utility program provides

d) the ability to introduce Frame Erasure channel impairment.

e) the ability to verify use of half-rate or lesser frame rate during dim-and-burst and packet level signaling

f) the ability to measure the Average Data Rate from an encoded packet file

A log output of `fersig29` provides detail on the ADR performance of the preceding encoder. In these applications, the utility is invoked as in following examples for 3% FER, and 1% signaling:

```
fersig29 -c EVRC-NW -e fer_3%.bin infile outfile

fersig29 -c EVRC-NW -s dim_1%.bin -e fer_3%.bin infile outfile
```

3.4.2.2 Channel Error and Signaling Masks

These binary Frame Error Rate and Signaling masks (source level and packet level) (1 byte of either 0 or 1 per frame) are used with the `fersig29` channel-impairment and inter-working simulation functions for the various conditions:

```
fer_3%.bin

dim_1%.bin

dim_1%_pls.bin
```

3.4.2.3 EVRC-NW Interworking Function (IWF)

The software “Evrc_nw_iwf.cc” can be compiled to yield a simulation utility `Evrc_nw_iwf` with usage defined as:

```
Evrc_nw_iwf –s signaling_mask_file –i encoded_packet_file –o dimmed_packet_file
```

where `Evrc_nw_iwf` converts full-rate frames in the input “encoded_packet_file” to half-rate frames at packet-level (that is using a simple scaling down of the packet instead of a complicated transcoding method).

3.4.2.4 P.341 Tx Filter

The software utility “p341_tx.c” can be compiled to yield a Tx filtering utility `p341_tx` with usage defined as:

```
p341_tx input-file-name output-file-name
```

where p341_tx is the 3GPP2 Tx filter compliant to ITU-T P.341. Figure 3.4.2.4-1 shows the frequency response of “p341_tx” filter. Also shown in this figure is the response of the ITU-T P.341 STL-2000 filter implementation, as well as the transmit masks for the ITU-T P.341/P.311 and the wideband transmit response from Table 9 in the 3GPP electro-acoustics specification [21]. From this figure, it can be seen that the STL-2000 filter response (in red) does not meet the frequency response of the 3GPP electro-acoustics specification, while the p341_tx filter response (in green) meets both the P.341/P.311 masks as well as the 3GPP electro-acoustics specification mask.

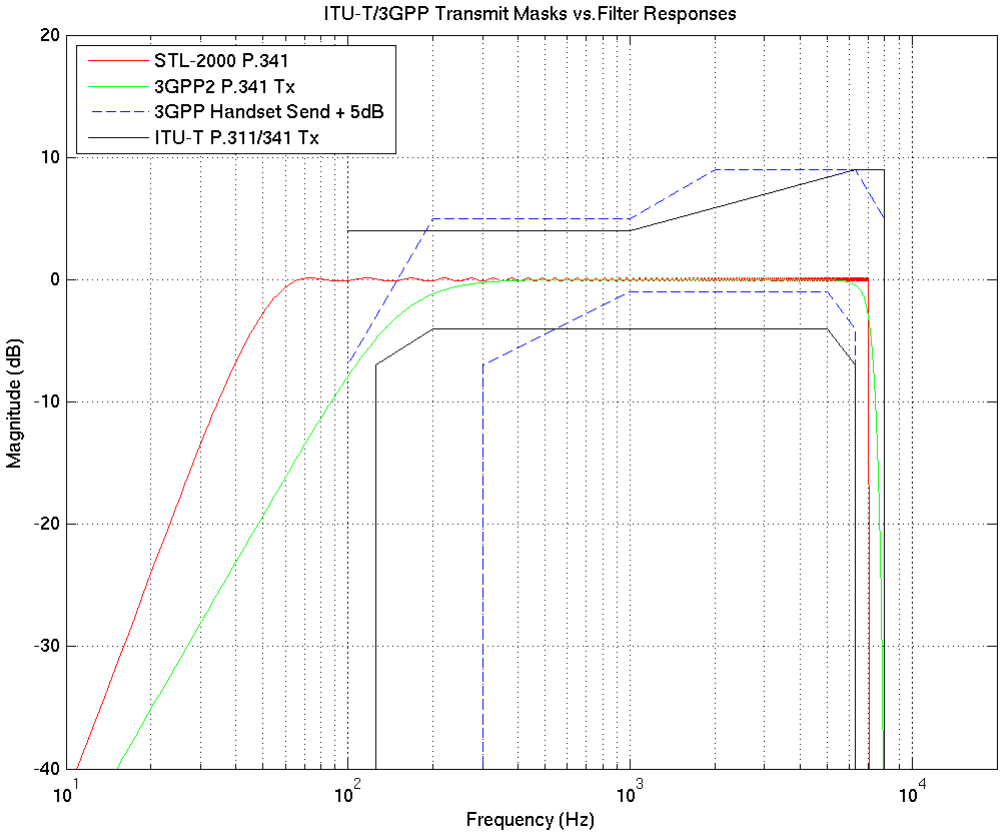


Figure 3.4.2.4-1 SO 73 ITU-T P.311/P.341 Transmit Mask and Filter responses

3.4.3 Master Codec for SO 73

This section describes the C simulation of the speech codec specified by [1]. The master codec C simulation used for verifying the performance of a non-bit-exact EVRC-NW implementation shall be the floating-point master C simulation included in the associated Software Distribution [1a].

3.4.3.1 Compiling the Master Codec Simulation

The source code for floating-point simulation can be compiled using the GNU G++ compiler and make utility.

A G++ compatible makefile has been included in the appropriate sub-directory in [1a]. Typing "make" this directory will compile and link the code and create the executable file called Evrc_nw (Evrc_nw.exe on Win32 systems), which will be placed in the same directory. The included makefile may require some user modification for a particular hardware platform and/or operating system.

3.4.3.2 Running the Master Codec Simulation

The EVRC-NW floating point executable (Evrc_nw) files use command line arguments to receive all information regarding input and output files and various parameters used during execution.

Executing “Evrc_nw” with no command line arguments will display a brief description of the required and optional command line arguments. The options are described below:

-i infn (required)	Specifies the name of the input speech file, or the name of the input packet file if only decoding is being performed (see the -d option below).
-o outf (required)	Specifies the name of the output speech file, or the name of the output packet file if only encoding is being performed (see the -e option below).
-d	Instructs the simulation to perform only the decoding function. The input file must contain packets of compressed data.
-e	Instructs the simulation to perform only the encoding function. The output file will contain packets of compressed data.
	If neither the -d or the -e option is invoked, the coder performs both the encoding and decoding functions by default.
-M max	Sets the maximum allowable data rate to max, where max is element of 4, 3, 2, 1, using the codes specified in the first column of Table 3.4.3.3-1.
-m min	Sets the minimum allowable data rate to min, where min is element of 4, 3, 2, 1, using the codes specified in the first column of Table 3.4.3.3-1.
	If neither the -M nor -m option is invoked, the coder allows the data rate to vary between Rate 1 and Rate 1/8.

3.4.3.3 File Formats for SO 73

Files of speech contain 2’s complement 16-bit samples with the least significant byte first. The packet file contains twelve 16-bit words with the low byte ordered first followed by the high byte.

The first word in the packet contains the data rate while the remaining 11 words contain the encoded speech data packed in accordance with the tables specified in [1]. The packet file value for each data rate is shown in Table 3.4.3.3-1.

Table 3.4.3.3-1 Packet File Structure from Master Codec/Channel Error Model

Value in Packet File	Rate	Data Bits per Frame
4 = 0x0004	1	171

Value in Packet File	Rate	Data Bits per Frame
3 = 0x0003	1/2	80
2 = 0x0001	1/4	40
1 = 0x0001	1/8	16
255 = 0x00FF	1/8 (Non-Critical)	16
0 = 0x0000	Blank	0
14 = 0x000e	Erasure	0

Unused bits are set to 0. For example, in a Rate 1/8 frame, the packet file will contain the word 0x0100 (byte-swapped 0x0001) followed by one 16-bit word containing the 16 data bits for the frame (in byte-swapped form), followed by ten 16-bit words containing all zero bits.

3.4.4 Fixed-Point Bit-Exact Codec for SO 73

This section describes the C simulation of the speech codec specified by [1]. The speech codec C simulation is based on finite precision, fixed-point arithmetic operations and is recommended to be used as a reference codec to verify the performance of a bit-exact EVRC-NW implementation of the fixed-point C simulation of a test codec. The bit-exact EVRC-NW codec, along with the appropriate test vectors to verify the bit-exactness performance, are included in the associated Software Distribution.

3.4.4.1 Fixed-Point Codec Program Files

This section describes the C program files which are provided in the associated software distribution for this document.

3.4.4.2 Compiling the Fixed-Point Codec Simulation

The source code for the fixed-point codec simulation has been written in C++ and can be compiled using any general purpose compiler such as the GNU G++ compiler and make utility.

Two GCC compatible makefiles have been included in the /build directory. Typing “make” in the /build directory will compile and link the code and create the executable file called Evrc_nw_fx (Evrc_nw_fx.exe on Win32 systems), which will be placed in the /build directory. The included makefiles may require some user modification for a particular hardware platform and/or operating system.

3.4.4.3 Running the Fixed-Point Codec Simulation

The EVRC-NW executable files use command line arguments to receive all information regarding input and output files and various parameters used during execution.

Executing Evrc_nw_fx with no command line arguments will display a brief description of the required and optional command line arguments. The options are described below:

-i infn (required) Specifies the name of the input speech file, or the name of the input packet file if only decoding is being performed (see the -d option below).

-o outf (required)	Specifies the name of the output speech file, or the name of the output packet file if only encoding is being performed (see the -e option below).
-d	Instructs the simulation to perform only the decoding function. The input file must contain packets of compressed data.
-e	Instructs the simulation to perform only the encoding function. The output file will contain packets of compressed data.
	If neither the -d or the -e option is invoked, the coder performs both the encoding and decoding functions by default.
-M max	Sets the maximum allowable data rate to max, where max is element of 4, 3, 2, 1, using the codes specified in the first column of Table 3.4.3.3-1.
-m min	Sets the minimum allowable data rate to min, where min is element of 4, 3, 2, 1, using the codes specified in the first column of Table 3.4.3.3-1.
	If neither the -M nor -m option is invoked, the coder allows the data rate to vary between Rate 1 and Rate 1/8.
	In addition, if max ≠ min, the data rate varies between max and min using the same rate decision algorithm, where the data rate is set to max if the selected data rate is >= max, and the data rate is set to min if the selected data rate is <= min.

3.4.4.4 File Formats

Files of speech contain 2's complement 16-bit samples with the least significant byte first. The packet file contains twelve 16-bit words with the low byte ordered first followed by the high byte.

The first word in the packet contains the data rate while the remaining 11 words contain the encoded speech data packed in accordance with the tables specified in [1]. The packet file value for each data rate is shown in Table 3.4.3.3-1. Unused bits are set to 0. For example, in a Rate 1/8 frame, the packet file will contain the word 0x0100 (byte-swapped 0x0001) followed by one 16-bit word containing the 16 data bits for the frame (in byte-swapped form), followed by ten 16-bit words containing all zero bits.

3.4.4.5 Verifying Bit-Exact Performance of the Fixed-Point Test Codec

This section outlines the methodology of verifying whether a Fixed-point Test codec is bit-exact to the Fixed point reference software. The purpose of this testing is to evaluate the bit-exactness of the test codec under a variety of conditions which may occur. To accomplish this, suites of test vectors have been designed to test for bit-exactness of the Test Codec under a variety of conditions depending on a number of parameters. These conditions include channel impairments, audio background noise, and different input levels.

Figure 3.4.4.5-1 illustrates a decision tree to arrive at the suite of test-vectors that are needed to demonstrate Minimum Performance Spec compliance through bit-exactness of a Test implementation of SO 73 for different profiles of equipments that support SO 73.

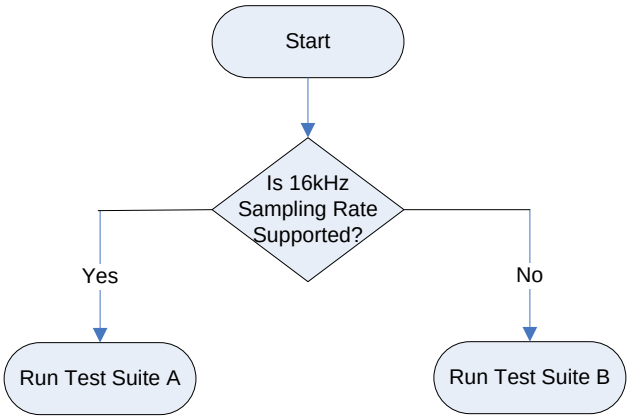


Figure 3.4.4.5-1 SO 73 Fixed-point bit-exact test suite decision flowchart

An implementation may support SO 73 only for 8 kHz sample rate input/output (for example, a Base-station transcoder or a Media Gateway). An implementation may support SO 73 for both 16 kHz and 8 kHz sample rate (for example, a mobile station that supports wideband electro-acoustics).

The main parameter in the decision tree is 16 kHz support in the implementation.

Depending on the implementation profile of the Device under test, one of 2 possible Test Suites are to be used to demonstrate SO 73 compliance. These 2 test suites named Test suites A and B, and the individual input test vectors comprising the Test suites are highlighted in Table 3.4.4.5-1.

Table 3.4.4.5-1 Test Suites of input test vectors for SO 73 compliance

Test Suites	Directory containing input test vectors
A	/so73/testvec/source/suiteA
B	/so73/testvec/source/suiteB

Files in the /so73/testvec/ directory are provided for the purpose of qualifying a test codec as bit-exact, and conform to the file-naming convention described in Section 2.2.4:

The /so73/testvec directory is divided into 2 subdirectories: /so73/testvec/source, and /so73/testvec/fixed.

The /so73/testvec/source directory contains input source files as well as packet files injected with frame erasures. The /so73/testvec/fixed directory contains files processed with the EVRC-NW fixed point reference software. The files in these directories are the reference files for bit-exact compliance. A test codec is bit-exact if it can reproduce all of the reference files in the /so73/testvec/fixed directory exactly. The outputs of the encoder and decoder of the test codec are to be obtained for the conditions given below in Table 3.4.4.5-2 – Table 3.4.4.5-5.The processing steps for these conditions are illustrated in Section 6.

Table 3.4.4.5-2 SO 73 Encoder Suite A Bit-exact Test Conditions

Input File	Operating Point	Condition	Reference packet files for bit-exact compliance
src.s22	EVRC-NW, operating point 0, 16 kHz sampling	Nominal, -22 dB	evrc_nw_op0.p22
src.s12	EVRC-NW, operating point 0, 16 kHz sampling	High, -12 dB	evrc_nw_op0.p12
src.s32	EVRC-NW, operating point 0, 16 kHz sampling	Low, -32 dB, 1% d&b	evrc_nw_op0.dim_1%.p32
src.c10	EVRC-NW, operating point 0, 16 kHz sampling	Nominal, -22 dB, 10 dB car noise	evrc_nw_op0.pc1
src.c20	EVRC-NW, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB car noise	evrc_nw_op0.pc2
src.s15	EVRC-NW, operating point 0, 16 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_nw_op0.ps
src.b20	EVRC-NW, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_nw_op0.pb
src.s22.8k	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB	evrc_nw_op1.p22
src.s12.8k	EVRC-NW, operating point 1, 8 kHz sampling	High, -12 dB	evrc_nw_op1.p12
src.s32.8k	EVRC-NW, operating point 1, 8 kHz sampling	Low, -32 dB	evrc_nw_op1.p32
src.s22.8k	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 1% d&b	evrc_nw_op1.dim_1%.p22
src.s22.8k	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB	evrc_nw_op6.p22
src.s12.8k	EVRC-NW, operating point 6, 8 kHz sampling	High, -12 dB	evrc_nw_op6.p12
src.s32.8k	EVRC-NW, operating point 6, 8 kHz sampling	Low, -32 dB	evrc_nw_op6.p32
src.s22.8k	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 1% d&b	evrc_nw_op6.dim_1%.p22
src.s22.8k	EVRC-NW, operating point 7, 8 kHz sampling	Nominal, -22 dB	evrc_nw_op7.p22
src.c15.8k	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op1.pc
src.s15.8k	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_nw_op1.ps
src.b20.8k	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_nw_op1.pb
src.c15.8k	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op6.pc
src.s15.8k	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_nw_op6.ps

Input File	Operating Point	Condition	Reference packet files for bit-exact compliance
src.b20.8k	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_nw_op6.pb
src.c15.8k	EVRC-NW, operating point 7, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op7.pc

Note: 9.3 kbps mode is generated using anchor operating point 0 and 5.8 kbps mode is generated using anchor operating point 2.

Table 3.4.4.5-3 SO 73 Suite A Decoder Bit-exact Test Conditions

Input Packet File	Operating Point	Condition	Reference output speech files for bit-exact compliance
evrc_nw_op0.fer_3%.p22	EVRC-NW, operating point 0, 16 kHz sampling	Nominal, -22 dB, 3% FER	evrc_nw_op0.fer_3%.o22
evrc_nw_op0.fer_1%.pls_1%.p22	EVRC-NW, operating point 0, 16 kHz sampling	Nominal, -22 dB, 1% FER, 1% pls	evrc_nw_op0. fer_1%.pls_1%.o22
evrc_nw_op0.p12	EVRC-NW, operating point 0, 16 kHz sampling	High, -12 dB	evrc_nw_op0.o12
evrc_nw_op0.dim_1%.p32	EVRC-NW, operating point 0, 16 kHz sampling	Low, -32 dB, 1% d&B	evrc_nw_op0.dim_1%.o32
evrc_nw_op0..pc1	EVRC-NW, operating point 0, 16 kHz sampling	Nominal, -22 dB, 10 dB car noise	evrc_nw_op0.oc1
evrc_nw_op0._fer_3%.pc2	EVRC-NW, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB car noise, fer_3%	evrc_nw_op0.fer_3%.oc2
evrc_nw_op0.ps	EVRC-NW, operating point 0, 16 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_nw_op0.os
evrc_nw_op0.pb	EVRC-NW, operating point 0, 16 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_nw_op0.ob
evrc_nw_op0.fer_3%.pm	EVRC-NW, operating point 0, 16 kHz sampling	Generic audio signal, fer_3%	evrc_nw_op0.fer_3%.om
evrc_nw_op1.fer_3%.p22	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, FER 3%	evrc_nw_op1.fer_3%.o22.8k
evrc_nw_op1.p12	EVRC-NW, operating point 1, 8 kHz sampling	High, -12 dB	evrc_nw_op1.o12.8k
evrc_nw_op1.p32	EVRC-NW, operating point 1, 8 kHz sampling	Low, -32 dB	evrc_nw_op1.o32.8k
evrc_nw_op7.p22	EVRC-NW, operating point 7, 8 kHz sampling	Nominal, -22 dB	evrc_nw_op7.o22.8k
evrc_nw_op1.dim_1%.pls_1%.p22	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 1% d&b, 1% pls	evrc_nw_op1.dim_1%.pls_1%.o22.8k

Input Packet File	Operating Point	Condition	Reference output speech files for bit-exact compliance
evrc_nw_op1.pc	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op1.oc.8k
evrc_nw_op7.pc	EVRC-NW, operating point 7, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op7.oc.8k
evrc_nw_op1.ps	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_nw_op1.os.8k
evrc_nw_op1.fer_2%.pb	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 15 dB babble noise, 2% FER	evrc_nw_op1.fer_2%.ob.8k
evrc_nw_op6.fer_3%.p22	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, FER 3%	evrc_nw_op6.fer_3%.o22.8k
evrc_nw_op6.p12	EVRC-NW, operating point 6, 8 kHz sampling	High, -12 dB	evrc_nw_op6.o12.8k
evrc_nw_op6.p32	EVRC-NW, operating point 6, 8 kHz sampling	Low, -32 dB	evrc_nw_op6.o32.8k
evrc_nw_op6.dim_1%.pls_1%.p22	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 1% d&b, 1% pls	evrc_nw_op6.dim_1%.pls_1%.o22.8k
evrc_nw_op6.pc	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op6.oc.8k
evrc_nw_op6.ps	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_nw_op6.os.8k
evrc_nw_op6.fer_2%.pb	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 15 dB babble noise, 2% FER	evrc_nw_op6.fer_2%.ob.8k

Table 3.4.4.5-4 SO 73 Encoder Suite B Bit-exact Test Conditions

Input File	Operating Point	Condition	Reference packet files for bit-exact compliance
src.s22.8k	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB	evrc_nw_op1.p22
src.s12.8k	EVRC-NW, operating point 1, 8 kHz sampling	High, -12 dB	evrc_nw_op1.p12
src.s32.8k	EVRC-NW, operating point 1, 8 kHz sampling	Low, -32 dB	evrc_nw_op1.p32
src.s22.8k	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 1% d&b	evrc_nw_op1.dim_1%.p22
src.s22.8k	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB	evrc_nw_op6.p22
src.s12.8k	EVRC-NW, operating point 6, 8 kHz sampling	High, -12 dB	evrc_nw_op6.p12

Input File	Operating Point	Condition	Reference packet files for bit-exact compliance
src.s32.8k	EVRC-NW, operating point 6, 8 kHz sampling	Low, -32 dB	evrc_nw_op6.p32
src.s22.8k	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 1% d&b	evrc_nw_op6.dim_1%.p22
src.s22.8k	EVRC-NW, operating point 7, 8 kHz sampling	Nominal, -22 dB	evrc_nw_op7.p22
src.c15.8k	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op1.pc
src.s15.8k	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_nw_op1.ps
src.b20.8k	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_nw_op1.pb
src.c15.8k	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op6.pc
src.s15.8k	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_nw_op6.ps
src.b20.8k	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 20 dB babble noise	evrc_nw_op6.pb
src.c15.8k	EVRC-NW, operating point 7, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op7.pc

Table 3.4.4.5-5 SO 73 Suite B Decoder Bit-exact Test Conditions

Input Packet File	Operating Point	Condition	Reference output speech files for bit-exact compliance
evrc_nw_op0.fer_3%.p22	EVRC-NW, operating point 0, 8 kHz sampling	Nominal, -22 dB, 3% FER	evrc_nw_op0.fer_3%.o22.8k
evrc_nw_op0.p12	EVRC-NW, operating point 0, 8 kHz sampling	High, -12 dB	evrc_nw_op0.o12.8k
evrc_nw_op0.p32	EVRC-NW, operating point 0, 8 kHz sampling	Low, -32 dB	evrc_nw_op0.o32.8k
evrc_nw_op0.dim_1%.pls_1%.p22	EVRC-NW, operating point 0, 8 kHz sampling	Nominal, -22 dB, 1% d&b, 1% pls	evrc_nw_op0.dim_1%.pls_1%.o22.8k
evrc_nw_op0.pc	EVRC-NW, operating point 0, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op0.oc.8k

Input Packet File	Operating Point	Condition	Reference output speech files for bit-exact compliance
evrc_nw_op0.ps	EVRC-NW, operating point 0, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_nw_op0.os.8k
evrc_nw_op0.fer_2%.pb	EVRC-NW, operating point 0, 8 kHz sampling	Nominal, -22 dB, 20 dB babble noise, 2% FER	evrc_nw_op0.fer_2%.ob.8k
evrc_nw_op0.fer_3%.pm	EVRC-NW, operating point 0, 8 kHz sampling	Generic audio signal, fer_3%	evrc_nw_op0.fer_3%.om.8k
evrc_nw_op1.fer_3%.p22	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, FER 3%	evrc_nw_op1.fer_3%.o22.8k
evrc_nw_op1.p12	EVRC-NW, operating point 1, 8 kHz sampling	High, -12 dB	evrc_nw_op1.o12.8k
evrc_nw_op1.p32	EVRC-NW, operating point 1, 8 kHz sampling	Low, -32 dB	evrc_nw_op1.o32.8k
evrc_nw_op6.fer_3%.p22	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, FER 3%	evrc_nw_op6.fer_3%.o22.8k
evrc_nw_op6.p12	EVRC-NW, operating point 6, 8 kHz sampling	High, -12 dB	evrc_nw_op6.o12.8k
evrc_nw_op6.p32	EVRC-NW, operating point 6, 8 kHz sampling	Low, -32 dB	evrc_nw_op6.o32.8k
evrc_nw_op7.p22	EVRC-NW, operating point 7, 8 kHz sampling	Nominal, -22 dB	evrc_nw_op7.o22.8k
evrc_nw_op1.dim_1%.pls_1%.p22	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 1% d&b, 1% pls	evrc_nw_op1.dim_1%.pls_1%.o22.8k
evrc_nw_op1.pc	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op1.oc.8k
evrc_nw_op6.dim_1%.pls_1%.p22	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 1% d&b, 1% pls	evrc_nw_op6.dim_1%.pls_1%.o22.8k
evrc_nw_op6.pc	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op6.oc.8k
evrc_nw_op7.pc	EVRC-NW, operating point 7, 8 kHz sampling	Nominal, -22 dB, 15 dB car noise	evrc_nw_op7.oc.8k
evrc_nw_op1.ps	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_nw_op1.os.8k
evrc_nw_op1.fer_2%.pb	EVRC-NW, operating point 1, 8 kHz sampling	Nominal, -22 dB, 15 dB babble noise	evrc_nw_op1.fer_2%.ob.8k
evrc_nw_op6.ps	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 15 dB street noise	evrc_nw_op6.os.8k
evrc_nw_op6.fer_2%.pb	EVRC-NW, operating point 6, 8 kHz sampling	Nominal, -22 dB, 15 dB babble noise	evrc_nw_op6.fer_2%.ob.8k

1

This page intentionally left blank.

4 CONTENTS OF SOFTWARE DISTRIBUTION

The source code for the master codec, fixed-point bit-exact codec and software tools, as well as the material needed to perform the objective and subjective tests described in this document are provided within an associated Software Distribution. The directory structure of the Software Distribution is represented in Table 4-1, Table 4-2, Table 4-3, and Table 4-4. Table 4-1 contains a brief description of the Software Distribution for the EVRC-A MPS, Table 4-2 contains a brief description of the Software Distribution for the EVRC-B MPS, Table 4-3 contains a brief description of the Software Distribution for the EVRC-WB MPS, and Table 4-4 contains a brief description of the Software Distribution for the EVRC-NW MPS. The prime sub-directories of these distributions are /so3, /so68, /so70, or /so73, respectively. These tables contain brief descriptions of the contents of these directories as well as cross-references to the sections of this document in which they are described in detail.

Table 4-1 Description of EVRC-A Software Distribution Contents

Directory	Description	References
/so3/simul/fixed	source code for the bit-exact fixed-point code	3.1.4
/so3/subjctv	Speech and other material necessary to perform Subjective Experiments I and II.	2.1.3, 2.1.4, 2.1.5
/so3/objctv	Speech material necessary to perform the Average Data Rate.	2.1.1
/so3/cal	Output level calibration file for listening tests.	3.1.1.2
/so3/tools	Source code for the software tools.	3.1.2
/so3/testvec	Test vectors for verifying bit-exact EVRC implementations.	3.1.4.6

Table 4-2 Description of EVRC-B Software Distribution Contents

Directory	Description	References
/so68/EVRCB_FX	source code for the bit-exact fixed-point code	3.2.4
/so68/subjctv	Speech and other material necessary to perform Subjective Experiments I and II.	2.2.1.1, 2.2.3, 2.2.4, 2.2.5
/so68/cal	Output level calibration file for listening tests.	3.2.1.2
/so68/tools	Source code for the software tools.	3.2.2
/so68/testvec	Test vectors for verifying bit-exact EVRC implementations.	3.2.4.5

Table 4-3 Description of EVRC-WB Software Distribution Contents

Directory	Description	References
/so70_73/EVRCWB_NW_FX	source code for the bit-exact fixed-point code	3.3.4
/so70/subjctv	Speech and other material necessary to perform subjective experiments.	2.3.1.1, 2.3.3, 2.3.4, 2.3.5
/so70/cal	Output level calibration file for listening tests.	3.3.1.2
/so70/tools	Source code for the software tools.	3.3.2
/so70/testvec	Test vectors for verifying bit-exact EVRC implementations.	3.3.4.5

Table 4-4 Description of EVRC-NW Software Distribution Contents

Directory	Description	References
/so70_73/EVRCWB_NW_FX	source code for the bit-exact fixed-point code	3.3.4
/so73/subjctv	Speech and other material necessary to perform subjective experiments.	2.3.1.1, 2.4.3, 2.3.4, 2.3.5
/so73/cal	Output level calibration file for listening tests.	3.3.1.2
/so73/tools	Source code for the software tools.	3.3.2
/so73/testvec	Test vectors for verifying bit-exact EVRC implementations.	3.3.4.5

This page intentionally left blank

5 DUNNETT'S TEST

Most of the MPS statistical tests for SO 68, SO 70, and SO 73 compliance involve multiple Test Encoder/Decoder Combinations (E/DC) and a single Reference E/DC. The appropriate analysis for the statistical tests involved in the EVRC-B MPS and EVRC-WB MPS test is Dunnett's Test [20]. Dunnett's Test is a special case of the more general Post Hoc Multiple Means Test, where multiple treatment means are statistically compared to a common control mean. In the case of the MPS tests, the treatments are the three Test E/DC's [M-T, T-M, T-T] and the control is the Reference E/DC [M-M].

Dunnett's Test is conducted in two stages. The first stage involves an Analysis of Variance (ANOVA) for the effects of *E/DC* x *Subjects*, where the *E/DC* factor includes the four E/DC's (three test E/DC's plus the Reference E/DC) and the *Subjects* factor includes the 32 subjects involved in the subjective test[9]. If the F-ratio for the *E/DC* effect is significant (i.e., $p < .05$) then there is significant variation among the scores for the E/DC's and the Dunnett's test proceeds to the second stage of the process. An F-ratio that is not significant indicates that there is no significant variation among the Test and Reference E/DC's. A non-significant F-ratio indicates that the means for all four E/DC's are statistically equivalent therefore all Test E/DC's are "not worse than" the Reference E/DC and all pass the MPS.

In the second stage of Dunnett's Test, each of the Test E/DC means is compared statistically to the Reference E/DC mean and the mean difference is evaluated for significance. The three statistical tests use a common estimate of the Standard Error of the Mean Difference (SE_{MD}) derived from the Error Mean Square from the ANOVA.

5.1 Stage 1 – Analysis of Variance

Table 5.1-1 shows the generalized Variance Source Table for the stage-1 ANOVA's involved in the Dunnett's Tests. The *Error* Sum of Squares (SoS) in the ANOVA is the residual SoS after removal of the systematic effects due to the *E/DC* and the *Subjects* factors.

9 The scores for each subject are average values over talkers.

Table 5.1-1 Variance Source Table for the ANOVA

Source	Degrees of Freedom (df)	Sum of Squares (SoS)	Mean Square (MS)	F-Ratio
E/DC	# E/DC's [*c*] - 1	$SoS_c = \sum_c s\left(\overline{X}_{.c} - \overline{X}_{cs}\right)^2$	$MS_c = SoS_c / df_c$	MS_c / MS_r
Subjects	# Subjects [*s*] –1	$SoS_s = \sum_s c\left(\overline{X}_{s.} - \overline{X}_{cs}\right)^2$		
Residual	$df_t - df_c - df_s$	$SoS_r = SoS_t – SoS_c - SoS_s$	$MS_r = SoS_r / df_r$	
Total	(c x s) - 1	$SoS_t = \sum_c \sum_s \left(x - \overline{X}_{cs}\right)^2$		

5.2 Stage 2 – Dunnett's Multiple Means Test — Test CC's vs. the Reference CC

In Stage 2 of the Dunnett's Test, the Mean score for each of the Test E/DC's ($\overline{X}_c$) is compared statistically to the Mean for the reference codec ($\overline{X}_{ref}$) as shown in Equation 5.2-1. The value for the Standard Error of the Mean Difference (SE_{MD}) is computed using the estimate of Mean Square Error (MS_E) derived from the Stage-1 ANOVA. The equation for computing SE_{MD} is shown in Equation 5.2-2, where MS_E is the Residual Mean Square from the ANOVA — MS_R in Table 5.1-1.

$$D_c = \left(\overline{X}_c - \overline{X}_{ref}\right) \Big/ SE_{MD} \quad (5.2\text{-}1)$$

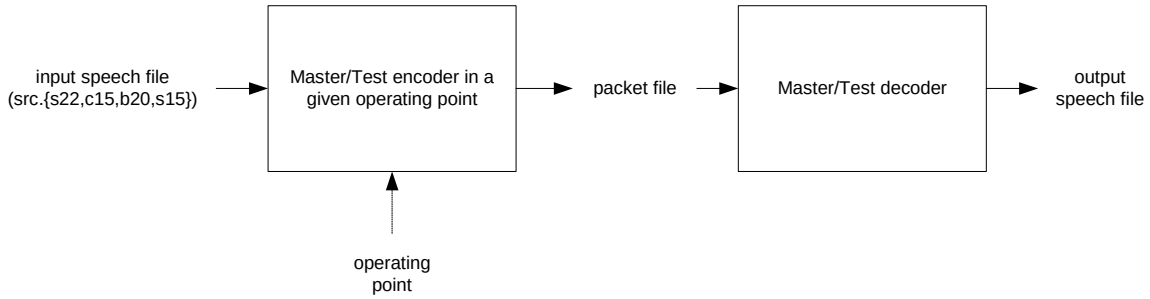
$$SE_{MD} = \sqrt{\left(2 \times MS_E\right) / \left(\#subjects\right)} \quad (5.2\text{-}2)$$

For each Test CC, the computed value of D_C is compared to critical values of the Dunnett's statistic, where the parameters are:

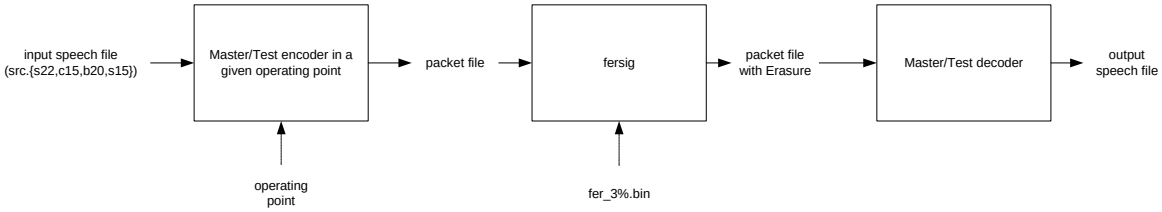
- criterion probability — p<.05
- total number of CC's (4)
- degrees of freedom for the MS_E (df = 93)
- Dunnett = 2.09

6 PROCESSING BLOCKS FOR SO 68, SO 70, AND SO 73

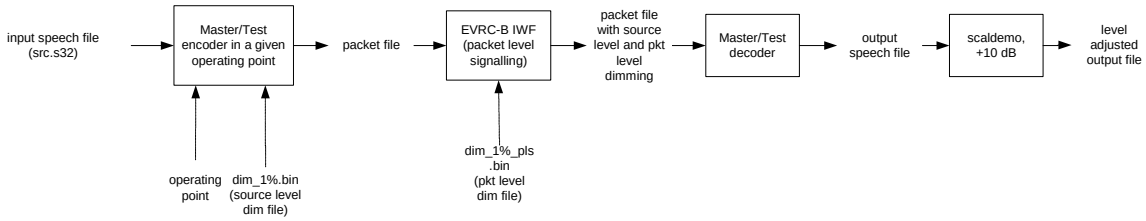
6.1 Nominal Level, and Noise Processing



6.2 FER Processing



6.3 Low-level, and Signaling Processing


input speech file
(src.s32)
Master/Test encoder in a given operating point
packet file
EVRC-B IWF (packet level signalling)
packet file with source level and pkt level dimming
Master/Test decoder
output speech file
scaldemo, +10 dB
level adjusted output file
operating point
dim_1%.bin (source level dim file)
dim_1%_pls .bin (pkt level dim file)

6.4 High level Processing